Selection Issues in the Analysis of Wages and in the Analysis of Electoral Outcomes

Dissertation
zur Erlangung des akademischen Grades
doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)

eingereicht an der
Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

Dipl.-Vw. Jakob de Lazzer

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Wirtschaftswissenschaftlichen Fakultät:
Prof. Dr. Daniel Klapper

Gutachter:
Prof. Bernd Fitzenberger, Ph.D.
Prof. Dr. Alexandra Spitz-Oener

Datum des Kolloquiums: 25.04.2019

Abstract

This thesis comprises four essays which study effects of non-random selection from different perspectives. Two essays examine the phenomenon of rising inequality of wages in Germany. The first one analyzes the role of changes in labor force composition for the development of wage inequality among full-time workers. Of particular interest are the effects of increasingly common episodes of temporary part-time work and nonemployment among full-time workers. Such episodes, along with changes in composition with respect to age and education, have contributed substantially to the rise in wage inequality.

Expanding on these results, the second essay studies the effects of declining unemployment on inequality of full-time wages. Selection between unemployment and full-time work is often determined by unobserved factors. Changing selection over time between those two labor market conditions could lead to increasingly diverse wages, particularly if sinking unemployment implies an influx of negatively selected workers into employment. However, results from a selection corrected quantile regression approach show that changing selection with respect to unobservables is not a contributor to the rise in wage inequality.

The third essay in this thesis studies non-monotonic selection in regression discontinuity designs (RDD). When similar numbers of individual select into and out of treatment simultaneously, the identifying assumption of the RDD can be violated. The essay describes the selection mechanisms and demonstrates the practical relevance of non-monotonic sorting in RDD applications, using election data. It then suggests an enhancement to the standard specification tests for RDDs, which can be used to detect non-monotonic sorting in advance.

Expanding the analysis of electoral results, the fourth essay studies voter's valuation of candidate gender. Despite significant improvement over the last decades, women are still heavily underrepresented in most countries' parliaments. This essay examines whether the presence of profession information coupled with voter preferences for stereotypical male occupations may explain part of this gap. The analysis is conducted as a field experiment built into an exit-poll of voters in Germany in 2014. Participants faced different versions of a hypothetical open list of candidates who exogenously varied their gender and profession. Comparing the voting behavior across different treatments shows a vote share bonus for women in the absence of profession information. Once voters know the profession of candidates, however, this changes towards a small edge for male candidates.

Zusammenfassung

Diese Dissertation umfasst vier Aufsätze welche die Effekte von nicht-zufälliger Selektion aus unterschiedlichen Perspektiven betrachten. Zwei der Aufsätze behandeln das Phänomen gestiegener Lohnungleichheit in Deutschland. Im ersten davon wird analysiert welche Rolle die geänderte Zusammensetzung der Erwerbsbevölkerung für die Entwicklung der Lohnungleichheit unter Vollzeit-Beschäftigten spielt. Vollzeit-Erwerbstätige weisen zunehmend Episoden von Teilzeitarbeit und Erwerbsunterbrechungen in ihren Biographien auf. Zusammen mit Veränderungen in der Alters- und Bildungsstruktur der Beschäftigten hat dies maßgeblich zum Anstieg der Lohnungleichheit beigetragen.

Aufbauend auf diesen Ergebnissen betrachtet der zweite Aufsatz die Effekte von gesunkener Arbeitslosigkeit auf Lohnungleichheit. Selektion in Vollzeit oder Arbeitslosigkeit erfolgt oft auf Basis unbeobachteter Faktoren. Wenn sich die Selektion in diese zwei Erwerbszustände über die Zeit ändert, kann dies zu steigender Lohnungleichheit führen. Dies ist insbesondere dann der Fall wenn Personen neu in Beschäftigung kommen, die eine Negativauswahl der Erwerbsbevölkerung sind. Jedoch zeigt sich, als Resultat einer Analyse mit selektionskorrigierten Quantilsregressionen, dass die veränderte Selektion nach unbeobachteten Faktoren nicht zum Anstieg der Lohnungleichheit beigetragen hat.

Im dritten Aufsatz dieser Dissertation geht es um nicht-monotone Selektion bei Regression Discontinuity Designs (RDD). Die Annahmen, auf welchen unverzerrte RDDs beruhen, können verletzt sein wenn sich sowohl Individuen in die Maßnahmengruppe hinein, als auch aus ihr heraus selektieren. Der Aufsatz beschreibt diesen Selektionsmechanismus und zeigt seine praktische Relevanz für RDDs auf Basis von Wahldaten. Weiterhin wird ein Spezifikationstest vorgestellt um nicht-monotone Selektion im Vorfeld der Analyse zu erkennen.

Als Weiterführung der Analyse von Wahldaten untersucht der vierte Aufsatz Wählerpräferenzen für das Geschlecht politischer Kandidaten. Trotz erheblicher Fortschritte in den letzten Jahrzehnten sind Frauen nach wie vor in Parlamenten unterrepräsentiert. In diesem Aufsatz wird analysiert ob das Zusammenspiel von Berufsinformationen und Geschlecht der Kandidaten die geringeren Repräsentation von Frauen erklären kann. Dazu wurde ein Feldexperiment durchgeführt, bei dem deutsche Wähler im Jahr 2014 beim Verlassen von Wahlbüros befragt wurden. Die Teilnehmer wählten Kandidaten aus verschiedenen Versionen einer synthetischen Parteiliste, bei der die Geschlechter und Berufe der Kandidaten zwischen den Versionen exogen variierten. Beim Vergleich des Wahlverhaltens über die verschiedenen Versionen zeigt sich dass weibliche Kandidaten einen Stimmvorteil genießen solange keine Berufsinformationen angegeben sind. Sobald jedoch die Berufe der Kandidaten bekannt sind kehrt sich dies in einen Stimmvorteil für männliche Kandidaten um.

Acknowledgments

Many people have supported my work over the course of this dissertation. First of all, I would like to thank my supervisor Bernd Fitzenberger for his continuous support. I have benefited from many discussions with him, both in his role as supervisor and as co-author, which has helped me improve this thesis. Working with him has granted me a lot of insight into the academic world. Moreover, I am particularly grateful for his support in the acquisition of funding for the entire duration of the dissertation.

I would also like to thank Alexandra Spitz-Oener, my second supervisor, for her support and for particularly helpful comments on my fourth research project presented in chapter 5.

I thank my co-author on the same project, Christoph Sajons, for the exceptionally fruitful collaboration on our joint research. I extend my thanks to Martin Biewen, who co-authored on the research paper presented in chapter 2.

I gracefully acknowledge financial support by the Deutsche Forschungsgemeinschaft through the priority program “The German labor market in a globalized world”.

Additionally, I thank my colleagues Julian Emmler, Stephan Martin, Marina Furdas and especially Markus Zimmermann. The pleasant work environment which we shared was very helpful for my progress on this thesis. I would also like to extend my thanks to the many people whose comments, advice or logistical support have helped me with this thesis. They are, in no particular order: Benjamin Bruns, Alfred Garloff, Blaise Melly, Lukas Mödl, Jan Nimczik, Florian Weiss, Stefan Krüger, Rainer Voß, Cornelia Hupfer and Lukas Mergele.

My research has profited from helpful comments and in-depth discussions from participants at the EALE 2018, the SSSI Bonn 2016, numerous SPP network meetings and conferences, the yearly meetings of the VFS and the internal research seminars of the chair of econometrics.

Last but not least, I thank my parents for their immeasurable support and generosity, and my friends for helping me keep my spirits high during stressful phases. Finally, my special thanks go to Elisabeth for her patience and encouragement over the course of the dissertation.

Legal Note

The four main chapters of the dissertation (Chapters 2-5) are self-contained and can be read independently.

Chapter 2 is joint work with Bernd Fitzenberger and Martin Biewen and has been published as a research article in the peer-reviewed *IZA Journal of Labor Economics* in 2018.

Chapter 3 is joint work with Bernd Fitzenberger.

Chapter 4 is single authored.

Chapter 5 is joint work with Christoph Sajons.

Contents

Chapter 1

Introduction

The study of non-random selection is central to the social sciences and selection issues are a topic of intensive study in economics. Selection simply means that the units of observation are non-randomly assigned to groups or categories, and that assignment is determined by a selection rule. It can be the result of the data collection process, or it can arise when agents self-select into specific options. For instance, individuals work in occupations in which their skills are most highly rewarded, workers only enter the labor market if they gain utility from employment and politicians are voted into office according to voter's perception of their quality. In the applied literature in economics, selectivity is an important issue in any application and has been so for a long time. As early as 1951, Roy notes that "those persons engaged in a particular occupation tend to be selected in a purposive manner from the working population as a whole." This distorted representation of the actual population is at the core of the selection problem (Heckman 1990b). Issues arise when comparing outcomes across groups with differing rules governing selection into them. However, the process of selection itself can be informative of economic conditions and therefore be an outcome of interest (Heckman and Vytlacil, 2007b). In this dissertation, I highlight selection issues from different perspectives. I analyze the economic effects of selection on inequality of wages and I study selective voter preferences for female political candidates. Methodologically, I also examine the challenges which selection issues present for the application of two different econometric approaches: quantile regressions and regression discontinuity designs.

Selection issues in economic research can be divided into two main strands. The first category is selection with respect to observable characteristics. These might be selective changes over time in the composition of the sample under analysis, e.g. the labor force, full-time workers (Lemieux 2006), the urban population (Moretti 2013) or migrants (Borjas 1987). These changes are responsible for shifts in outcomes over time and have to be taken into account when analyzing developments. They can also themselves be the object of analysis (Heckman and Vytlacil 2007a).

In chapter 2, my co-authors and I focus on the effect of changes in the observed composition of the labor force on the distribution of wages in Germany. Germany has experienced a dramatic rise in inequality of wages between the early 1990s and 2010 (see for instance Antonczyk et al. 2018, Fitzenberger 2012). Greater disparity in wages has informed the political debate and might have contributed to the introduction of a minimum wage in Germany in 2015 (compare Brenke and Müller 2013). Multiple factors have added to the increase in inequality: Skill biased technological change (Spitz-Oener 2006), changing labor market institutions (Antonczyk et al. 2010, Felbermayr et al. 2014b) and assortative matching of firms and workers (Card et al. 2013). The labor force has also been subject to a large shift towards higher education, aging of the population, rising female labor force participation (Dustmann et al. 2009). Put differently, selection into the labor force, with respect to observed individual characteristics, has changed over time. Re-examining the development of the wage distribution in Germany, we use administrative panel data to investigate the role of composition changes in education, age, individual labor market histories, occupations and industry for the rise in inequality. We show that these changes in composition have contributed substantially to the rise in wage inequality and introduce a novel explanatory factor: An increase in the in the incidence of employment interruptions and temporary part-time work among full-time workers has had a large, positive effect on the spread of the wage distribution. We observe that between 1985 and 2010, full-timer workers have become more likely to have experienced part-time work or employment interruptions at some previous point in their labor market history. This change towards "patchwork" labor market histories can be observed for both males and females.

We decompose the rise in inequality into components due to changes in each characteristics group, using an inverse probability reweighting approach in the spirit of DiNardo et al. (1996). Our results show that changes in observables account for a large part of the rise in wage inequality, and that the growing importance of employment interruptions and temporary part-time episodes play an important role for wage inequality among full-time workers. For males, we find that 43 to 53 percent, depending on the baseyear, of the rise in wage inequality between 1985 and 2010 can be explained by compositional effects of observables. Of those, 14 to 17 percentage points are due to increasing employment interruptions and temporary part-time work. For females the importance of composition changes is even higher, ranging between 64 and 78 percent, with 17 to 18 percentage points due to changing labor market histories. These results are also policy relevant because both changes in the age/education structure and in labor market histories are observable and to a certain extent predictable. We therefore expect the positive trends in patchwork labor market histories to directly change the wage distribution in the future.

The second category of selection issues is those of selection with respect to unobservable factors. In a labor market context, certain individual traits like determination, social skills or adaptability are hard to measure and are usually not observed, particularly in administrative data. Additionally, the outcome of interest may only be observed for a selective sample which differs from the population in terms of unobservables. If selection is not random, the observed distribution of the outcome will not be representative of the population. As a consequence, estimated effects of covariates on the outcome will generally be biased. This problem has sparked an extensive literature about approaches for selection corrected models. For use in regression analysis, Heckman (1979) pioneered the use of control function estimators to correct estimated coefficients for bias due to selection with respect to unobservables and a range of other approaches have since been established (for instance Heckman 1990a, Puhani 2000 and Buchinsky 1998).

In chapter 3, my coauthor and I study the wage effect of non-random selection into full-time work in Germany, taking into account changes in unobservables. The German labor market has experienced large shifts between full-time work and unemployment since the mid-90s (Ljungqvist and Sargent 1998). Simultaneously, the wage distribution has widened. Understanding the influence of changing selection on wage spreads is therefore important for evaluating the economic effects of rising wage inequality. If falling unemployment draws people into employment who represent a negative selection of all workers, the rise in inequality will be overestimated. In this case, increasing wage inequality can be considered the sign of a positive development because it implies that individuals who previously did not have employment are entering work.

We therefore study two aspects of selection into employment. First, we quantify the magnitude of inequality which would be observed if all unemployed were working full time, using exogenous labor supply shocks as instruments for selection. Second, we consider the counterfactual development of inequality if the pattern of selective movement between unemployment and full-time work had not changed since the mid-90s. Hereby we use a large administrative dataset of west-German male workers and restrict the duration of unemployment to a maximum of one year. Our econometric approach relies on quantile regressions which we correct for selection into full-time work. It is an enhancement of Buchinsky's (2001) approach and takes into account the critique by Huber and Melly (2015). It is part of a small but econometrically advanced literature about correcting estimated distributions for selection (see Buchinsky 2001, Arellano and Bonhomme 2017, D'Haultfoeuille et al. 2014).

The most striking result from our analysis is that changing selection over time with respect to unobervables is not driving the increase in wage inequality. For the subgroup of medium educated workers, changing unobservables did not have a significant influence on within-group inequality. For the low educated, wage inequality would actually have increased if selection with respect to unobservables had stayed as it was in 1995. This implies that full-time workers have become less heterogeneous with respect to unobservable characteristics. These results have some political relevance, because they show that rising wage inequality cannot be explained by an influx of previously unemployed individuals into full-time work. Rising inequality in wages is therefore a sign of greater disparity in observable worker skills and the returns to those skills. We also find that in a given year there is positive selection into employment with respect to unobservables, especially in the lower parts of the wage distribution. Which means that the employed are a positive selection of the labor force, even after conditioning on education and labor market history.

Selection is not just an issue in labor market research and a challenge for the application of quantile regression methods. In the context of field experiments, non-random selection into treatment or control groups will lead to bias in estimated treatment effects. Selection, sometimes also in the form of noncompliance or attrition, is a common feature of both field and social experiments (Heckman and Smith 1995). Any experimental analysis therefore has to take into account the potential for selection issues (Heckman and Vytlacil 2007a; 2007b).

In chapter 4, I highlight a specific kind of selection issues with respect to unobservables in regression discontinuity designs (RDD). These research designs are a type of natural experiment which exploits variation in treatment status caused by a discontinuity in a continuous individual variable. Although first proposed by Thistlethwaite and Campbell in 1960, the approach has only seen widespread application starting in the early 2000s (compare Lee 2001, Hahn et al. 2001, Dinardo and Lee 2004, among others). If the running variable passes a specific threshold, treatment status changes. As long as individuals can't exert perfect control over their realization of the running variable, treatment assignment close to the threshold is as good as randomized. This allows for the estimation of local average treatment effects (LATE) by fitting a flexible model on both sides of the threshold. If, however, individuals can precisely manipulate their value of the running variable and sort themselves just above or below the threshold, the identifying assumption of the design is violated and the LATE is biased. This is a special case of selection, which can be driven by either observable or unobservable characteristics of the individuals. In the literature it is standard practice to detect sorting issues by checking for continuity of control variables across the threshold (compare Imbens and Lemieux 2008). Thus, selection with respect to observable characteristics in the RDD setting can be detected. However, selection with respect to unobservables will remain hidden. A different specification test, proposed byMcCrary (2008), is based on the idea that sorting at the threshold should create a discontinuity in the density of observations at this point. In chapter 4, I point out that, while the McCrary-test is very good at detecting monotonic sorting, it can't detect non-monotonic sorting issues. Monotonic sorting takes place if all individuals have homogeneous preferences with respect to treatment status. Non-monotonic sorting occurs when some individuals non-randomly sort into treatment while others sort out of treatment, which can also invalidate the identifying assumption of the RDD.

I show that this kind of sorting is a practical concern for RDD applications, using data from the analysis of incumbency advantages in U.S. House elections by Lee (2001). The application exploits the supposedly random outcomes of extremely close elections to provide a causal estimate of the vote share advantage for incumbent candidates. There is however substantial evidence that close elections are not as random as assumed and that non-monotonic sorting takes place in them (Alvarez and Hall 2006, Caro 1990 and Caughey and Sekhon 2011). I demonstrate that the McCrary test cannot detect sorting issues in this application, and suggest a modification to the test, based on subsample-testing of groups with monotonic selection preferences. Using this approach, I provide evidence that significant sorting takes place in this application. Some candidates in U.S. house races are able to precisely sort themselves above the vote share threshold needed for winning, while simultaneously sorting their competitor slightly below this threshold. Because candidates of both parties are equally capable of sorting, voluntary selection into treatment is masked by involuntary selection out of treatment.

While chapters 2-4 cover issues which arise under selective movement into treatment or labor market status, in chapter 5, the process of selection itself stands at the core of the analysis. Using data from an election experiment built into an exit-poll of voters in Germany in 2014, my co-author and I examine voter demand for female political candidates.

By performing an experiment which randomly assigns treatment to individuals, we avoid many types of selection issues typical of non-experimental data and are able to precisely identify the causal effects of candidate gender. One caveat of this approach is that we cannot entirely rule out self-selection into the experiment. However, we minimize selection issues with regards to voting participation by drawing respondents only from those individuals who actually went to the polls. This also ensures that our respondents are in the right set of mind to cast their vote as they would do in real elections.

Despite significant improvement over the last decades, women are still heavily underrepresented in most countries' parliaments (CAWP 2018, IPU 2018). And while there is ample evidence that supply decisions of female candidates contributes to this disparity, the literature has not reached a conclusion about the effects of voter demand for female candidates (compare Esteve-Volart and Bagues 2012, Black and Erickson 2003 Giger et al. 2014, among others). Our results suggest that voter demand for female candidates depends strongly on the availability of information about the profession of the candidate. This might also help explain the wide range of results, found in the literature, regarding voter demand for female candidates.

In our experiment, participants faced different versions of a hypothetical open list of candidates who exogenously varied their gender and profession. We also included a ballot version without profession information, in order to separate the effects of gender from those of profession. Each candidate appears on different ballot versions with a male-dominated, a female-dominated and a gender-neutral profession, respectively.

Comparing the voting behavior across different treatments, we do not find direct discrimination against women in the absence of profession information. Female candidates even enjoy a substantial electoral advantage, driven by female voters. Female candidates are ~33 percent more likely to receive the vote than their male counterparts. Once voters know the profession of candidates, however, this vote share advantage vanishes and would even change towards a small edge for male candidates if professions were realistically distributed. This effect is caused by the fact that voter preferences for female candidates are replaced by voter preferences for candidates with stereotypical combinations of gender and profession. Men working in a male-dominated professions enjoy a larger vote share bonus than that of any other group, while women working in female-dominated professions gain a smaller vote share bonus.

Our results are policy relevant for local or list elections, in which professions are important for characterizing candidates. Since we find no evidence that voters are generally biased against women as candidates, our results imply that policy measures which aim to improve female representation in parliaments can align with voter preferences. Gender stereotypes play an important role in determining vote share and carry some relevance for electoral campaigns. Candidates in gender-typical professions might want to emphasize their occupation, while candidates with atypical combinations of gender and profession might focus on other characteristics.

Chapter 2

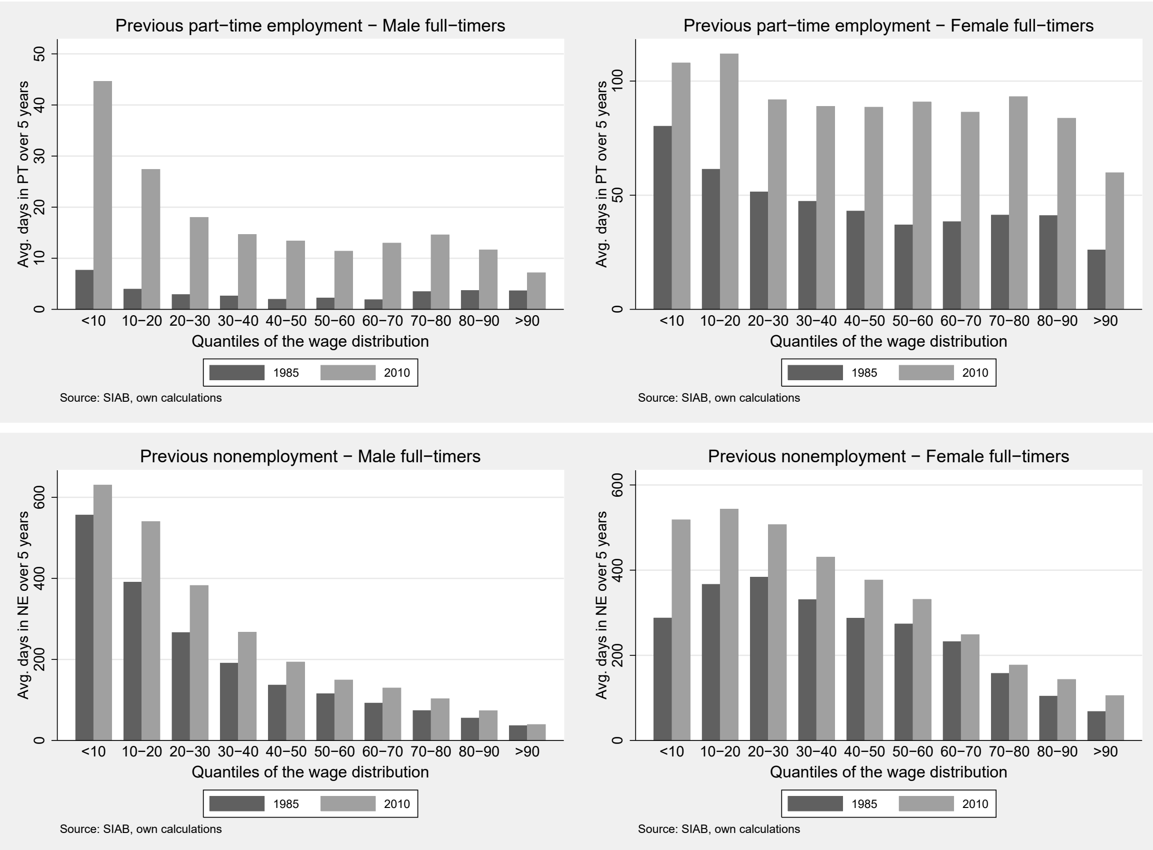
The Role of Employment Interruptions and Part-time Work for the Rise in Wage Inequality

2.1 Introduction

The incidence of employment interruptions and temporary part-time work has grown strongly, raising concerns about the stability of employment and low wages among part-time workers (OECD 2010). Less known is that the incidence of previous part-time work and employment interruptions has also grown among full-time workers. However, employment interruptions and part-time experience may be associated with lower future wages due to lower human capital accumulation, negative signalling effects, or lower labor force attachment (Arulampalam 2001, Blundell et al. 2016, Heckman 1981, Paul 2016). The literature on the rise in wage inequality among full-time workers has so far not taken this into account. This is the first study to examine the impact of changes in recent labor market histories on the rise in wage inequality. Re-examining the development of the wage distribution in Germany, we use administrative panel data to investigate the role of composition changes, in particular changes in recent labor market experience, for the rise in wage inequality.[1] As the key novel aspects, our study accounts explicitly for previous part-time work and employment interruptions among full-time employees, and we extend the analysis to total employment.

[1]There is a large literature on the rise of wage inequality in Germany, see e.g. Dustmann et al. 2009 Dustmann et al. 2009, Antonczyk et al. 2010, Card et al. 2013 as well as the literature review in section 2.2.

Figure 2.1: Part-time employment and Nonemployment during previous five years in different parts of the full-time wage distribution



Note: Average number of days in part-time employment/nonemployment during the years 1980-1984 and 2005-2009, respectively, by decile of the full-time wage distribution in the years 1985 and 2010.

Motivating our analysis, figure 2.1 shows for the years 1985 and 2010 the number of days in part-time employment and nonemployment, respectively, during the previous five years by decile of the wage distribution. For full-timers both the incidence of previous part-time and nonemployment experience increased considerably between 1985 and 2010. Put differently, full-timers have over time become more likely to have experienced part-time work or employment interruptions in the past. The prevalence of previous part-time experience and nonemployment increases in the lower part of the full-time wage distribution, implying that among workers with particular low wages the share of workers, who have recently worked part-time or who have experienced nonemployment in the recent past, has grown over time. Figure 2.1 shows that nonemployment experience is more important than part-time experience, with male (female) full-timers in 2010 in the lowest decile having experienced an average of more than 600 (500) days of nonemployment and more than 40 (110) days of part-time employment during the time period 2005 to 2009. The evidence for part-time employment is consistent with studies showing that part-time work has increased strongly and that transitions between part-time and full-time work and employment interruptions have become more frequent (Tisch and Tophoven (2012) , Potrafke (2012), Tamm et al. (2017)). Below, we will also show evidence that the dispersion of nonemployment and part-time experience among full-timers has grown over time. There was a secular increase of unemployment in Germany from the 1980s until the mid 2000s. Afterwards, unemployment fell almost continuously until 2010 (SVR 2014). Our analysis will focus on long-term changes abstracting form cyclical variation in nonemployment and part-time experience among full-timers.[2]

There is ample evidence suggesting that episodes of part-time work or nonemployment have negative long-term impacts on the career path and therefore on future wages.[3] First, human capital accumulation slows down or there is even depreciation when workers interrupt their career or temporarily downgrade to part-time employment. Second, employment interruptions or part-time experience may lead to scarring effects leading to lower wage offers and poorer career possibilities upon re-employment. A third point is that lagged employment outcomes are indicators of permanent characteristics which drive employment and wages. Accordingly, periods of nonemployment or part-time employment in the past may indicate a lower labor force attachment - in addition to being a negative productivity signal. Lagged employment outcomes are unobserved in the cross-sectional data sets, typically used in the literature on wage inequality for most countries (see e.g. Acemoglu and Autor 2011 and the literature discussion in section 2.2).

For the aforementioned reasons, our paper investigates the role of employment interruptions and part-time employment in a statistical decomposition of the rise in wage inequality among full-time working employees. In light of the evidence in figure 2.1, the growing importance of part-time employment and nonemployment is likely to play an important role for the increase of lower tail wage inequality. The literature review in section 2.2 reveals that the studies on the rise of wage inequality have so far not taken into account the rise in previous

[2]There is a cyclical component in transitions from nonemployment and part-time employment to full-time employment. During an upswing (downturn), one would expect these to increase (fall). In a recent study, Borowczyk-Martins and Lalé (2016) show that for the UK and the U.S. transitions from part-time to full-time employment at the same employer are a major driver of the cyclical changes in part-time employment, growing (declining) during an upswing (downturn). Our analysis focuses on the long-term rise in the share of full-timers with nonemployment and part-time experience. As our empirical results show, this long-term rise dominates the cyclical variation.

[3]See e.g. Arulampalam (2001), Burda and Mertens (2001), Beblo et al. (2002), Manning and Petrongolo (2008), Edin and Gustavsson (2008), Schmieder et al. (2010), Edler et al. (2015), or Paul (2016).

nonemployment and part-time employment among full-timers. Furthermore, little attention has been paid to gender differences in the rise in wage inequality. For instance, negative long-term career effects of transition from full-time to part-time work for women after childbirth have been studied by Connolly and Gregory (2009) and Paul (2016). Fitzenberger et al. (2016) document that women in Germany, who had been working full-time before birth, take fairly long spells of maternity leave after child birth and often then return to part-time work.

Our paper makes the following contributions. First, in our decomposition of the rise in wage inequality among full-timers, we add the previous labor market history involving part-time and nonemployment experience. This plays an important role in explaining the rise in wage inequality both among males and females. At the same time, adding previous labor market history accounts for unobserved heterogeneity in employment decisions. As such, our analysis is of interest for all countries experiencing similiar labor market trends, because ours is the first study investigating the role of the rise in nonemployment and part-time employment in explaining the rise in wage inequality among full-time employees. As a related second contribution we estimate the effect of further observable characteristics to the increase in male and female wage inequality in Germany over the recent decades. Such a parallel analysis for Germany does not exist. Compositional changes in observable characteristics explain over 50 percent of the increase in male wage inequality and up to 80 percent of the increase in female wage inequality. To the best of our knowledge, the extremely strong role of composition effects for the rise of female wage inequality has not been recognized so far. Third, we estimate composition effects with regard to the counterfactual distribution of full-time wages for all employees, which confirms the robustness of our main findings. Furthermore, this shows that part-timers (especially female part-timers) represent a negative selection with respect to observable characteristics. Including part-timers into the analysis also speaks to the role of increasingly heterogeneous labor market histories for the rise in German wage inequality.

The remainder of this paper is structured as follows. Section 2.2 reviews the literature on the rise of wage inequality. Section 2.3 discusses the data used and presents first descriptive evidence. Section 2.4 discusses our findings. Section 2.5 concludes. The appendix provides more details and supplementary empirical results.

2.2 Literature review

Wage inequality has been increasing in many industrialized countries between the 1980s and the 2000s (see the comprehensive survey in Acemoglu and Autor 2011, or the literature discussion in Lemieux 2006, Autor et al. 2008, Dustmann et al. 2009). Many studies focus on the U.S, but the same mechanisms operating in the U.S. are also at play in other industrialized countries, including Germany. Skill-biased technical change (SBTC) is the most prominent explanation for the rise in wage inequality, predicting rising wage inequality across the entire wage distribution. This is consistent with the evidence for the U.S. for the 1980s but not for the 1990s, as in the 1990s inequality stopped to grow at the bottom of the wage distribution (Autor et al., 2008). Acemoglu and Autor (2011) take the latter as evidence for the task-based approach (see Autor et al. 2003) implying a falling demand for occupations with medium skill requirements (which are relatively more routine intensive and thus easier to substitute by technology) relative to both occupations with high or with low skill requirements, resulting in polarization of employment across occupations. The evidence regarding a polarization of wages across the wage distribution in the U.S. seems to be limited to the 1990s and a polarization of wages is not an unambiguous prediction of the task based approach (Autor, 2013). Some studies for the U.S. emphasize the role of changing labor market institutions such as de-unionization and falling real minimum wages (see also the discussion in Autor et al. 2003). DiNardo et al. (1996) show that the fall in unionization levels explains an important part of the rise in wage inequality during the 1980s.

In related work, Lemieux (2006) shows that changes in the composition of the workforce regarding education and experience explain a major part of the rise in wage inequality in the U.S. Also, Autor et al. (2008) find strong composition effects, especially for females, but focus on other explanations for the rise in wage inequality. Composition effects also affect residual wage inequality, i.e. the wage differences among employees with the same observable characteristics (DiNardo et al. 1996, Lemieux 2006). Altogether, this evidence motivated us to scrutinize the role of composition effects for the rise of wage inequality in Germany.

Wage inequality has been rising in West Germany [henceforth Germany] since the 1980s (Dustmann et al., 2009).[4] Until the mid 1990s the increase in wage dispersion among full-

[4] See also (in chronological order) Kohn 2006, Gernandt and Pfeiffer 2007, Antonczyk et al. 2010, Fitzenberger 2012, Card et al. 2013, Felbermayr et al. 2014a, Dustmann et al. 2014, Riphahn and Schnitzlein 2016, Möller 2016, and Antonczyk et al. 2018. Most recent studies are based on administrative employment records in the Sample of Integrated Employment Biographies (SIAB) – or on earlier versions of the same data source - as provided by the Research Data Center of the IAB and the Federal Employment Agency. Some studies use of the cross-sectional wage surveys in the German Structure of Earnings Survey (GSES) provided by the

timers was restricted to the top of the wage distribution, whereas wage inequality increased from mid 1990s onwards until 2004 across the entire distribution (Dustmann et al., 2009). The evidence until the mid 1990s is consistent with skill biased technological change and the hypothesis that labor market institutions such as unions and minimum wages prevented an rise in wage inequality at the bottom of the wage distribution before the mid 1990s, which resulted in rising unemployment among the low-skilled (Fitzenberger, 1999). Dustmann et al. (2009) show that changes in the composition of workers regarding age and education and the sizeable decline in coverage by collective bargaining both explain major components of the rise in wage inequality. At the same time, the study provides evidence for a polarization of employment as found previously for the U.S. (see also Antonczyk et al. 2018).

Antonczyk et al. (2009) and Antonczyk et al. (2010) find a strong increase of wage inequality between 1999/2001 and 2006. Changes in task assignments cannot explain this rise (Antonczyk et al., 2009). Accounting for coverage by collective bargaining, firm level characteristics, and personal characteristics, Antonczyk et al. (2010) show that the decline in coverage by collective bargaining does not explain the rise in wage inequality in the lower part of the wage distribution, when firm level characteristics are held constant. Most important are changes in the quantile regression coefficients of firm level variables (firm size, region, industry), which reflect a growing heterogeneity in firm level wage policies. The two studies differ regarding the contribution of changes in personal characteristics. Biewen and Seckler (2017) find that changes in union coverage and personal characteristics are most important for the rise in wage inequality between 1995 and 2010. Card et al. (2013) estimate person and firm fixed effects in wages. The study finds a growing heterogeneity of these fixed effects over time and increasing sorting of workers with high personal fixed effects into firms with high firm fixed effects. Both effects contribute strongly to the rise in wage inequality. Felbermayr et al. (2014a) find that the decline in coverage by collective bargaining is the most important explanation for the rise in wage inequality, while there is no important role for international trade. Our short survey of the literature shows that the literature has not yet reached a consensus on the mechanisms behind the rise in wage inequality in Germany

Research Data Center of the Statistical Offices, the Socio-Economic Panel (GSOEP) provided by DIW or the BIBB-IAB/Bibb-BAuA Labor Force Surveys (BLFS). While the SIAB data only involves earnings, the GSES, the GSOEP, and the BLFS allow for an analysis of hourly wages. Researchers using the SIAB data typically focus on full-time working employees. While the SIAB and the GSOEP provide panel data, the GSES data and the BLFS only involve repeated cross-sections every four to six years and the GSES surveys before 2010 only involve a subset of all industries and they lack very small firms. Compared to the GSOEP and the BLFS, the GSES and the SIAB provide much larger cross-sections on employees and wages. All four data sets document the rise in wage inequality since the mid 1990s, see Dustmann et al. (2009, SIAB), Fitzenberger (2012, SIAB and GSES), Antonczyk et al. (2009, BFLS), and Gernandt and Pfeiffer (2007, GSOEP).

until 2010.[5]

None of the aforementioned studies investigates to what extent the rise in interruptions of full-time work is driving the increase in wage inequality, although there is ample evidence of a negative effect of previous nonemployment and part-time experience on wages in full-time employment. Several mechanism may be at work. First, human capital accumulation slows down or there is even depreciation when workers stop working full-time (Beblo et al. 2002, Manning and Petrongolo 2008, Edin and Gustavsson 2008, Paul 2016). Employment interruptions due to displacement have been shown to negatively affect wages (Burda and Mertens 2001, Schmieder et al. 2010, Edler et al. 2015). After maternity leave, females often return to part-time employment, but may return to full-time work later on (Fitzenberger et al. 2016, Paul 2016). When a transition from nonemployment or part-time work back into full-time work involves a job change (no recall), this also implies a loss of job-specific human capital. Second, employment interruptions or part-time experience may lead to scarring effects, i.e. employers (rightly or wrongly) interpret previous non-fulltime employment as signal of low productivity or low labor force attachment leading to lower wage offers and poorer career possibilities upon re-employment (Ruhm 1991, Arulampalam 2001, Gregory and Jukes 2001). A third potential mechanism, similar to the second, is that lagged employment outcomes are indicators of permanent characteristics which drive employment and wages (Heckman, 1981). Accordingly, periods of nonemployment or part-time employment in the past may indicate a lower labor force attachment - in addition to being a negative productivity signal.

The literature on wage effects of temporary part-time work focuses primarily on women and maternal part-time. For females in the UK, Connolly and Gregory (2009) and Blundell et al. (2016) demonstrate that part-time employment in the past results in lower earnings trajectories, even when returning to full-time work. Connolly and Gregory (2009) also show that this holds for part-time episodes at the same employer. They point out that part-time work is often related to downgrading to less skilled tasks that persists if the individual later returns to full-time work. Controlling for selection on unobservables, Paul (2016) finds for Germany a substantial negative impact of part-time work and nonemployment episodes on future earnings of females in full-time work, with the effect being even stronger for nonemployment. While there is no detailed analysis of part-time effects among males

[5]The recent study by Möller (2016) shows that the rise in wage inequality stopped in 2010 based on a new release of the SIAB data. However, the comparison of the years before and after 2011 is plagued by a structural break in 2011 regarding the distinction between part-timers and full-timers. For both reasons, we abstain from analyzing the SIAB data after 2010 since our focus is on analyzing the rise in wage inequality.

available, the mechanisms of human capital depreciation and lack of further training which underly the wage effects of part-time work for female workers are likely to affect male workers in a similar way.

2.3 Data and descriptive evidence

Our analysis uses SIAB data involving a 2% sample of all dependent employees who are subject to social security contributions, i.e. excluding the self-employed and civil servants.[6] We study the period 1985 to 2010. Even though SIAB data are available for earlier years, we do not include them in our analysis because the rise of wage inequality across the entire distribution is only observed after the 1980s (Dustmann et al. 2009, Fitzenberger 2012). Since we may observe several employment spells of various lengths per individual in a given year, all observations are weighted with the share of days worked in a job in the respective year. The sampling weights calculated in this way reflect the relative importance of each wage observation.

We account for an individual's labor market history using four measures. The first two involve the number of days spent in full-time and in part-time employment during the last five years. The residual category is the number of days spent in nonemployment during the last five years, which may be times of unemployment, education, or any other type of nonemployment. In addition, we use two dummy variables, indicating whether a person had a full-time or a part-time spell at any point during the previous year. This information captures individual short-term employment dynamics. Wages are daily wages in Euros deflated by the CPI to 1990. Since we use administrative data on employment spells, the measures are very precise. Because the SIAB data do not involve hours worked, we follow the literature on wage inequality for Germany and use daily wages, representing an earnings measure. Our sample also includes individuals with part-time employment, but the wage data for part-timers are much more confounded by differences in hours of work than for full-timers.[7] Below, we also estimate the counterfactual distribution of full-time wages for total employment also including part-timers.

[6]This study uses the factually anonymous Sample of Integrated Labour Market Biographies (version 1975 – 2010). Data access was provided via a Scientific Use File supplied by the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB), see vom Berge et al. (2013) for a data documentation.

[7]We have calculated the standard deviation of hours of work for the years 1985 and 2010 based on the German Socioeconomic Panel (detailed results are available upon request). For part-timers, the standard deviation is two to three times higher than for full-timers.

Table 2.1: Variable Classification

Variable group	Short	Variable list
Education	Ed	3 categories (*ed*): University, Upper secondary High-School and/or Vocational Training, No/Other Degree
Experience	Ex	Potential experience (age - years of schooling - 6) (ex)
Labor market history	Hist	Number of days in full-time ($ft5$), or part-time ($pt5$) over the last 5 years. Indicators for: full-time job in previous year (ft), part-time job in previous year (pt)
Occupation	Occ	Job classification by KldB 2-digit levels (*occ*, 63 categories)
Industry	Ind	Industry classification by WZ93 (*sec*, 14 categories)

All wages above the contribution threshold are top-coded in the SIAB. The censoring threshold lies above the 85% wage quantile in every year. Therefore, we compare the 85/15, the 85/50 and the 50/15 quantile gaps in the wage distribution. In those cases, where we cannot restrict our analysis to values below the 85% quantile (in particular when analyzing developments in wage residuals), we impute wages above the threshold according to individual characteristics. Details of the imputation procedure can be found in the appendix, section 2.6.3. Unless noted otherwise, we restrict our analysis to individuals aged 20 to 60 years, in order to focus on the working age population.

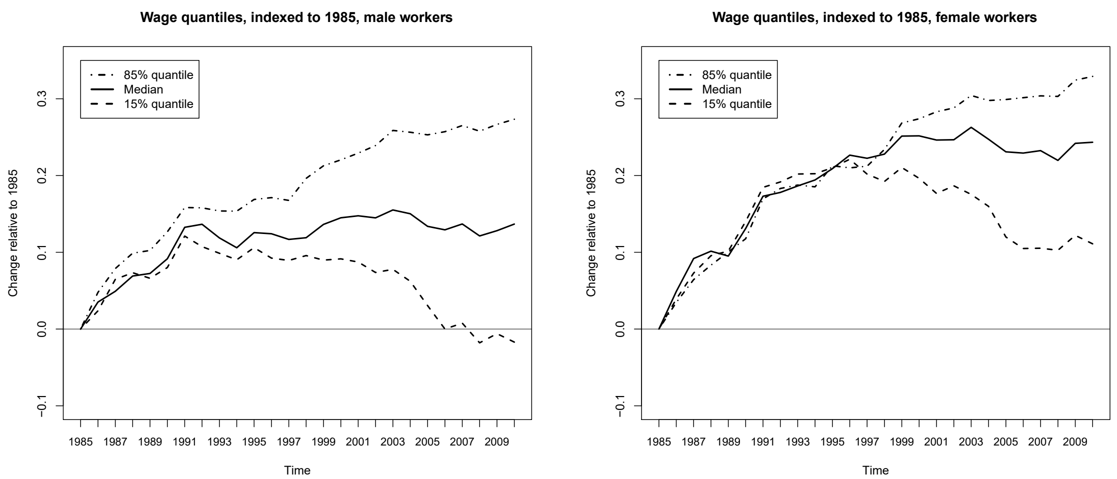
Table 2.1 lists the covariates used and Table A2.1 provides descriptive statistics for two sample years. We distinguish three education levels: University degree (including Universities of Applied Sciences), degree from Upper secondary school and/or Vocational Training, No/Other degree. We use 14 aggregated industries (German Industry Classification [WZ] 1993) and 63 aggregated occupations (2-digit level of the KldB ['Klassifikation der Berufe'] 1988). For interactions between industry and occupation, we aggregate occupations to the 1-digit level in order to avoid problems with empty cells in our logit regressions. The education variable is cleaned and interrupted measurements are imputed for consistency based on Fitzenberger et al. (2006).

2.3.1 Wage inequality

Figure 2.2 shows the development of log wage quantiles (cumulative changes) from 1985 onwards. Our primary measures of wage inequality are the gaps between the 85th, 50th and 15th percentiles of log wages. Until about 1991 the different wage quantiles move upward and largely in parallel. After 1991 median wages of male full-timers stagnate (recall that we analyze real wages). For female full-timers there is a continuous but decelerating rise until

2003 and a subsequent decline until 2008. For both genders we observe a widening of the wage distribution beginning just at the time when median wages start stagnating. Wages at the 85th percentile continue to increase, while wages at the 15th percentile decline. For males this decline is moderate until the early 2000s but accelerates afterwards. By 2010 male wages at the 15th percentile even lie below their 1985 level. For females we observe different developments of the three quantiles already in the late 1980s. However, inequality only increases in a more substantial way in the late 1990s, several years later than for males. After 1998 female median wages stagnate, while the 85th percentile rises and the 15th percentile declines rapidly. The corresponding trends in inequality as measured by the 85/50 and 50/15 gaps are depicted by the solid lines in figures A2.9 to A2.12.

Figure 2.2: Wage quantiles relative to levels of 1985



Note: Wage quantiles among full-time workers in West-Germany, normalized to the levels of 1985. Source: SIAB, own calculations.

2.3.2 Labor market histories

Part-time work in Germany has grown substantially over the last decades (figure A2.1). While this may reflect secular trends in labor market participation, part of the increase is linked to political reforms promoting part-time work. Over our observation period several changes in legislation focus on part-time work. In 1985 the German government enacted a law ('*Beschäftigungsförderungsgesetz*') which granted part-timers the same level of job protection as full-timers. This law increased the acceptance of part-time work on the side of trade unions and in the general population. In 2001 a law followed which made it easier for employees to enter voluntary part-time work ('*Teilzeit- und Befristungsgesetz*').

These changes in legislation had the effect of formally easing the transition between full-time, part-time and nonemployment. We observe that not only the yearly stock of part-time employees increased for both genders, but that the frequency of temporary part-time episodes for individuals currently working full-time increased as well (figure A2.3). Parallel to the rise of part-time work two changes in legislation between 1985 and 1998 ('*Beschäftigungsförderungsgesetz*', '*Arbeitsförderungs-Reformgesetz*') facilitated fixed-term contracts and temporary agency work.

Both the intensive and the extensive margin of labor market histories matter for current wages (Burda and Mertens 2001, Arulampalam 2001, Beblo et al. 2002, Manning and Petrongolo 2008, Edin and Gustavsson 2008, Schmieder et al. 2010, Edler et al. 2015, Paul 2016, Blundell et al. 2016). Returns to labor market experience are not uniform across jobs and type of work. Not only is experience in part-time work valued lower than that in full-time work, but part-time and nonemployment episodes slow down career progression and wage growth, see literature review in section 2.2.

Figure A2.3 shows increasing average lengths and also increasing variability of previous part-time episodes for men and females, both above and below the median of the respective wage distribution.[8] The mean and variance of number of days spent in part-time work during the last five years increases over time for those individuals who are in full-time jobs at the time of observation. Male full-timers experience a noticeable increase in past part-time episodes, although the total amount of the time previously spent in part-time is lower than for females. While the increase in prevalence of previous part-time for males is only slightly higher below than above the median, the increase in *variability* of previous part-time experience is considerably stronger below the median. This means that previous part-time episodes are increasingly concentrated on low-wage full-timers which may lead to rising lower-tail wage inequality.[9] For female full-timers we observe an increase in the length and variability of previous part-time work both above and below the median, and the overall levels are considerably higher than for males. Incidentally, the part-time experience of full-time females above the median of the distribution shows stronger cyclical variation compared to females below the median, whose part-time experience follows more of a secular upward trend. Note that the labor supply of females is known to be more elastic than that

[8] In order to clearly separate previous part-time and nonemployment during educational spells from those after having completed education, we also include the evidence for full-timers aged at 30 to 60 years old, see figures A2.15 and A2.16 in the appendix. For part-time experience, the trends are very similar for the 25 to 60 years old and the 30 to 60 years old.

[9] In table A2.8 in the appendix, we show that differences in means and variances below and above the median are highly statistically significant.

of men and that the part-time experience of females is often related to career interruptions after child birth (Blundell et al., 2016).

There are two further issues concerning temporary part-time episodes to be discussed. The first involves working time accounts which provided a buffer against the negative labor demand shock in Germany during the Great recession 2008/2009 (Burda and Hunt, 2011). The SIAB data do not record a variation in hours worked over a year in case of continuous employment at the same employer. In the case of working time accounts, the part-time/full-time classification is based upon agreed (contractual) hours of work. Furthermore, the data involve daily wages defined as total earnings over an employment spell (typically one year, when the worker is employed by the same employer for one calendar year) divided by the length of the employment spell in days. Specifically, working time accounts allow to vary the actual hours of work over a year but there is no variation in monthly earnings. Furthermore, on average over the employment spell the actual hours of work should correspond to the contractual ones. Note further that working time accounts did not play an important role before 2008 and that they show a strong cyclical variation. By contrast, our results below suggest an earlier timing of the distributional effects of previous part-time episodes, reflecting a long-term continuous trends which dominates the cyclical variation. The second issue concerns whether the part-time episodes in our data are with the same employer or with different employers. A recent study shows that a major part of the cyclical variation in part-time employment in the UK and the U.S. is accounted for by changes in transitions rates between part-time and full-time work at the same employer (Borowczyk-Martins and Lalé, 2016). We would expect wage penalties associated with previous part-time episodes to be larger if they occur across employers. Our data show that the vast majority (about 75-80%) of transitions from part-time to full-time employment involve a change of employers (see figure A2.2). We observe only a minor cyclical variation in the division of the part-time to full-time transitions within and between employers, which is unlikely to be of importance in explaining the continuous long-term rise in wage inequality (see decomposition results in section 2.4).

We now turn to the descriptive discussion of previous nonemployment episodes. Just as previous part-time experience, nonemployment has a sizeable negative impact on wages. Nonemployment may include all alternative activities such as education or child care or it may be due to involuntary displacement, unemployment or voluntary absence from the labor market. Such events may lead to human capital obsolescence, with the possible exception of educational spells, and therefore to a decline in wages (Burda and Mertens 2001, Schmieder et al. 2010, Edler et al. 2015). Figure A2.4 shows the average length and variability of time spent in nonemployment over the past 5 years. Both above and below the median, males and females exhibit increasing previous nonemployment experience. Cross-sectional variability only increases below the median, and there is a cyclical variation, which is stronger below the median. To investigate whether educational spells are driving our results, we reduce the sample to individuals age 30 years or above, for whom we assume that educational spells play a negligible role among nonemployment episodes, see figure A2.16 in the appendix. Above the median wage, the upward trend now disappears. By contrast, males and females below the median wage still exhibit increasing previous nonemployment experience together with increasing cross-sectional variability. Thus, previous nonemployment episodes are increasingly concentrated on individuals in the lower part of the wage distribution, a trend which may have a strong impact on lower-tail wage inequality. The differences between figures A2.4 and A2.16 reveals that educational spells are an important part of previous nonemployment episodes among younger workers.[10]

Irrespective of the type of previous nonemployment episodes, their incidence is higher than previous part-time employment, especially for males but also for females. Moreover, the associated wage losses are likely to be larger than those from part-time episodes (except for educational spells among younger workers, which may, however, be captured in our subsequent analysis by a higher education level). We therefore expect previous nonemployment episodes to have sizeable negative effects on wages, most likely raising lower-tail wage inequality.

2.3.3 Education, experience, industry, and occupation

In addition to the changes in recent labor market history, there have been strong changes in the distribution of education, work experience and industry structure. Figure A2.5 shows the percentage of workers in each education category. The share of workers without an educational degree has declined since the 1980s. This holds in particular for female workers, among whom the percentage of unskilled workers decreases from 32% in 1985 to 18% in 2010. We also observe an increase in the share of university graduates. Again, this is most pronounced for females, as the initial percentage of female university graduates is very small in 1985 but catches up to the male share by 2010. For the medium-skilled, i.e. workers with an upper secondary degree or a vocational degree, we observe a hump-shaped development. The share of medium-skilled increases during the late 1980s and the 1990s,

[10]Unfortunately, the SIAB data do not record whether a nonemployment episode is due to an educational spell. However, the data involve the educational degree as possible outcome of a previous nonemployment episode.

reaches its peak in the late 1990s, and declines in the 2000s, giving way to a rising share of university graduates.

The corresponding trends for the distribution of worker's potential experience are shown in figure A2.6. Between 1985 and 2010 the percentage of highly experienced workers with 27 or more years of potential experience increases, reflecting the aging of the population. The share of workers with medium levels of potential experience (between 14 and 26 years) follows a hump shaped trend. The percentage of older workers with 40 or more years of experience did not undergo major changes in our sample, even though the overall population aged considerably. The only major gender difference in potential work experience concerns the share of workers with low experience. Among males this share is never higher than 20% and it drops to 10% in the late 1990s. Starting at 30% in 1985, the initial share of young female workers is very high but converges to the low level for males in the late 1990s. After the catching-up process among females, the experience composition by gender has became very similar by 2010. Note that our experience measure is potential work experience which mainly reflects both workers' age and educational periods. In this way, we more clearly separate long-term trends in experience (population aging and educational periods) from the factors we intend to capture in our recent labor market histories (recent part-time and nonemployment episodes).

Figure A2.8 shows the development of industry shares for eight aggregated sectors. We observe some sectors with an almost constant share since the 1980s (i.e. transportation and trade), while others experiences strong changes. For males the largest changes are observed for the construction industry, the manufacturing sector for consumer goods, and the banking and insurance sector. The first two experience a massive decline, while the latter more than doubles its share between 1985 and 2010. Transport and communication as well as health and social services, show small increases, whereas the manufacturing sectors for vehicles and for machinery shrink slightly. The initial sector composition differs strongly by gender, but the dynamics of the different sectors are quite similar. In particular, manufacturing declines strongly, while banking and health services grows. The construction sector, which plays no important role for females, does not change in any substantial way.

Shifts between occupations are smaller than those between industries. Table A2.2 reports the five most frequent occupations in 1985 and 2010. Figure A2.7 in the appendix shows a continuous shift in the aggregate from manufacturing to service sector occupations. At the same time, there are fairly small changes in the distribution of the 63 two-digit occupations. Among males four out of five occupations are present in the top 5 in both years and their shares are similar. For females three out of five occupations remain in the top 5 in both years. Furthermore, the correlation coefficient between employment shares for the 63 two-digit occupations in 1985 and 2010 is .91 for males and .96 for females.

2.4 Empirical analysis

2.4.1 Estimation of counterfactuals

First, we analyze the impact of composition changes on wage inequality among full-timers accounting for the selection into full-time work based on observed worker characteristics. For the counterfactual analysis keeping characteristics constant over time we use the reweighting methodology introduced by DiNardo et al. (1996).[11] Then, we repeat the analysis for wage inequality for total employment in a similar way. We now provide a brief overview of what we do, full formal details can be found in the appendix.

We start to estimate the distribution of full-time wages which would result if the distribution of worker characteristics had not changed over time while the conditional wage distribution given worker characteristics changed over time as observed.[12] For example, we hold fixed the composition with respect to education and estimate as counterfactual by how much inequality would have risen if workers' education had not changed. We sequentially add groups of covariates in order to determine the incremental effect of a particular set of covariates. For example, in the situation in which we already leave education constant, we also fix workers' potential work experience in order to determine the incremental effect of experience to rising wage inequality. Our sequential conditioning scheme is such that we move from exogenous and predetermined characteristics towards characteristics that are the likely consequence of endogenous decisions of the individual. Altogether, we start with workers' education and sequentially add the factors potential work experience, recent labor market history as well as workers' occupation and industry (see table 2.1). As in Lemieux (2006), we also carry out our decomposition for *residual* wage inequality, i.e. wage inequality within groups of workers with identical observed characteristics.

In the second part of our analysis, we take the distribution of full-time wage earners, but reweight their characteristics to replicate the distribution of observed characteristics for

[11] This method has been applied, among others, by Lemieux (2006) and Dustmann et al. (2009). For an overview of alternative decomposition techniques, see Fortin et al. 2011

[12] Such an analysis ignores general equilibrium effects, i.e. changes in the conditional wage structure are assumed to be independent of changes in the work force composition.

total employment, i.e. including part-time workers. This estimates the counterfactual wage distribution that would result if all employed workers worked full-time. Contrasting this distribution with the wage distribution among full-timers allows one to gauge to which extent part-timers represent a positive or negative selection compared to full-timers. We repeat our sequential analysis of adding different groups of covariates for the reweighted sample representing total employment.

2.4.2 Wage inequality among full-timers

Starting with *male full-timers* we first analyze the effect of educational upgrading on male wage inequality. Figure A2.9 (left panel) shows the evolution of the quantile gaps in male wages between 1985 and 2010 under the assumption that the 1985 distribution of education is held fixed over time. It turns out that fixing education considerably reduces the increase in inequality, i.e. the observed educational upgrading contributes strongly to the observed rise in wage inequality. Table A2.4 shows that a share of 17.1% of the increase in overall inequality (as measured by the 85/15 quantile ratio) and 37.5% of the increase in the upper half of the distribution (as measured by the 50/15 quantile ratio) can be explained by changes in education, while these changes did not contribute to rising inequality at the bottom of the distribution (as measured by the 50/15 quantile ratio, see lower part of figure A2.9). This means that the compositional effects of the educational expansion mostly affected the upper part but not the lower part of the male wage distribution. The contribution of changes in education on residual wage inequality amounts to a moderate 7.1%, i.e. there is no strong shift towards groups of workers with above-average levels of within-group inequality. As a next step, figure A2.10 extends the reweighting procedure to include changes in work experience (in addition to changes in education). Based on the evidence shown in figure A2.10 (left panel) and table A2.4 (columns 4 to 6) the incremental contribution of work experience is very small.

In figure A2.11 we add changes in recent labor market histories to our reweighting procedure. This considerably changes the results, affecting in particular the bottom of the distribution. The incremental contribution amounts to 16.9% for overall wage inequality and to 19.2% for lower tail inequality (column 10 of table A2.4). This means that increasingly discontinuous labor market histories are important to explain the rise in lower-tail wage inequality. There was also a sizeable contribution to changes in residual wage inequality (10.7%), suggesting that changes in recent labor market histories were associated with shifts towards worker groups with higher levels of within-group inequality. Finally, figure A2.12 adds changes in occupations and industry structure. This also contributes to the general rise in male wage inequality (13.0% for overall wage inequality, 22.2% to inequality at the bottom, and 13.6% to residual wage inequality, see columns 11 to 13 of table A2.4).

Note that adding the stage (Occ+Ind) results in the cumulative effect of changing the joint distribution of all our covariates (Ed+Ex+Hist+Occ+Ind). As shown in column 12 of table A2.4, compositional changes explain more than half of the increase in male wage inequality over the period 1985 to 2010 (53.0% of overall wage inequality, 54.6%/51.5% at the top/bottom, 34.0% of residual wage inequality). Our results confirm the importance of compositional effects for male wage inequality changes also found by Dustmann et al. (2009) and Felbermayr et al. (2014a), but establish the contribution of the additional factor of changes in recent labor market histories. Note that the explanatory power of compositional changes is particularly high between 1985 and 1995 (holding characteristics fixed there is no increase in inequality at all, see left panel of figure A2.12), but became somewhat weaker from 1996 onwards. Similar to the findings for the U.S. (Lemieux, 2006), the total contribution of the compositional changes considered lies above 50%, which is quite high.

Next, we turn to results for *female full-timers*, see the right hand panels of figures A2.9 to A2.12. By contrast to the findings for males, figure A2.9 shows that the increase in female wage inequality remains largely unchanged, when holding constant the 1985 distribution of education.[13] Adding changes in potential work experience (which are mainly driven by age) yields a strong incremental contribution (35.1% to overall inequality, 30.4% to upper half inequality, and 38.2% to lower half inequality, see figure A2.10 and columns 5 to 7 of table A2.5). This also differs from the findings for males. In light of figure A2.6, the findings for females reflect that younger cohorts are much smaller compared to older ones (e.g. the share of females with 0 to 13 years of potential work experience dropped from 30% in 1985 to 10% in 2010). This leads to a rising share of older female full-timers with different wage levels and higher within-group inequality.

Adding recent labor market histories again explains a considerable, incremental share (18.6% for overall inequality and 17.1% for residual inequality, columnns 8 to 10 in table A2.5, see also figure A2.11 to the right). Thus, the impact of part-time episodes and labor market interruptions is similar for males and females. Finally, changes in occupations and industry structure have negligible effects on rising female wage inequality (columns 11 to 13 of table

[13]However, there is a slight difference with regard to the effect of female education when we take as the base year 2010 instead of 1985. This points to interaction effects. We carry out this reverse analysis in section section 2.6.4 in the appendix.

A2.5).[14]

Altogether, we find that compositional changes can account for an even larger share of the rise in female wage inequality than for males. Column 12 of table A2.5 shows that 63.6% of the increase in overall inequality, 61.9% of the increase in the upper part, and 64.8% of the increase in the lower part of the distribution can be accounted for by the compositional changes considered. The graph to the right in figure A2.12 implies that, during the period 1991 to 2001, female wage inequality would have fallen even in the absence of compositional changes. An important component has worked through composition changes regarding residual wages, i.e. shifts between groups of workers with different levels of within-group inequality (51.6% of the changes are accounted for by composition changes, see column 12 of table A2.5).

In the appendix, we carry out a robustness check of our analysis that reverses the roles of the base and target years (1985 vs. 2010). With few exceptions all our findings are robust to the choice of the base year (see appendix for details).

2.4.3 Counterfactual full-time wages for total employment

This section extends the analysis of full-wage wages to total employment, including those working part-time in the year of observation. As explained above, part-time wages are not comparable because we lack detailed information on hours worked in our data set. However, we do observe the personal characteristics of part-timers, which our analysis of composition effects includes. We consider the distribution of characteristics in the combined sample of full-timers and part-timers (‘total employment’), thereby estimating inequality of full-time wages among individuals who are currently employed.

This excercise will be informative in four ways. First, comparing the actual wage distribution of full-timers with the counterfactual wage distribution that assumes that both part-time and full-timers are paid full-time wages will be informative about whether part-timers are a positive or negative selection with respect to their characteristics (compared to full-timers). Second, examining the development of the counterfactual wage distribution for the total employment sample over time may serve as an estimate for composition effects on wage inequality in a wider population of part-time and full-timers, which we cannot examine

[14] It is not an error that quantile gaps for the overall distribution are unchanged up to the third digit in row 13 of table A2.5 when adding occupation and industry characteristics. This is due to the fact that daily wages are rounded to full Euros and quantiles only change if the change in counterfactual weights is large enough to move the quantile value to a different Euro integer.

directly given that comparable wage information for part-timers is missing. This also serves as a robustness checks of our above findings for full-timers. Third, the effect of selection into full-time work versus part-time work is mostly accounted for by controlling for the recent employment history. Fourth, we net out selective transitions between part-time and full-time work in our analysis of composition effects, in the sense that we measure composition effects net of such (often temporary) movements between part-time and full-time work.

We start with the estimated counterfactual trends in inequality of full-time wages in a sample sharing the composition of total employment (for a more detailed explanation, see section 2.6.4 in the appendix). Figure A2.13 shows the trend in wage inequality if full-timers shared the education composition of total employment. For male workers the differences between both distributions is very small in 1985. After 2000 we see a slight decline in the 15% quantile of the total employment distribution relative to the full-time distribution, which leads to slightly wider 50/15 and 85/15 quantile gaps. This suggests a negative selection into part-time work for men. However, the part-time share of male workers already starts rising in the early 1990s, while we only observe negative effects of selection into part-time a decade later. This implies that there is no negative selection associated with the initial expansion of part-time work. Also, for females the initial full-time and total employment distributions for females are quite similar, especially regarding the upper tail. However, the quantiles diverge quickly and by 1990 we see lower wages for the total employment sample over the entire distribution. This means that characteristics that were prevalent among part-time workers involve lower wage returns than those of full-timers, implying negative selection into part-time work. After 1990 the distributional gap between the full-time and the counterfactual total employment sample was almost constant, implying a stable positive selection into full-time work. The differences of the observed female full-time wage distribution in 2010 and the wage distribution for the counterfactual total emloyment sample are also shown in the right panel of figure A2.14 (bold vs. dashed line). Considering the total employment sample shifts the distribution to the left, i.e. the full-time sample is positively selected. The dotted lines in figure A2.14 represent the wage distributions that result when one further changes the characteristics to those of the total employment sample in 1985. This results in a considerable compression of the wage distribution. Again, changing characteristics contribute to rising inequality. Table A2.6 shows the contribution of composition changes for trends in full-time wage inequality in the total employment sample, which are broadly similar to the results for the male full-timers in table A2.4. In particular, there is an important role for composition changes regarding education (especially at the top) and labor market histories (especially at the bottom). Including part-timers into the analysis makes the contribution of labor market histories to rising inequality much more pronounced at the bottom of the distribution (38.6% in table A2.6 vs. 22.2% in A2.4). There is only a limited role for changes in occupations and industries. These conclusions are robust to reversing the base year, see tables A2.6 and A2.10 in the appendix. Table A2.7 shows the results for the female total employment sample. Despite the much higher part-time share in the female sample the results in table A2.7 are again quite similar to table A2.5 for female full-timers. There is a role for shifts in experience and recent labor market histories, while changes in education and occupations and industries do not contribute much. In table A2.11 in the appendix we reverse the base year. As in the female full-time sample this boosts the role of education changes (particularly at the top of the distribution) and leads to a number of smaller unsystematic changes that point to complex interaction effects of compositional and wage structure effects. Similar to males, extending the analysis to total employment for females also amplifies the importance of recent labor market histories for increasing wage inequality at the bottom of the distribution (20.5% vs. 28.6% in table A2.5 vs. table A2.7, and 11.9% vs. 22.7% in table A2.13 vs. table A2.11, column 10).

2.5 Conclusions

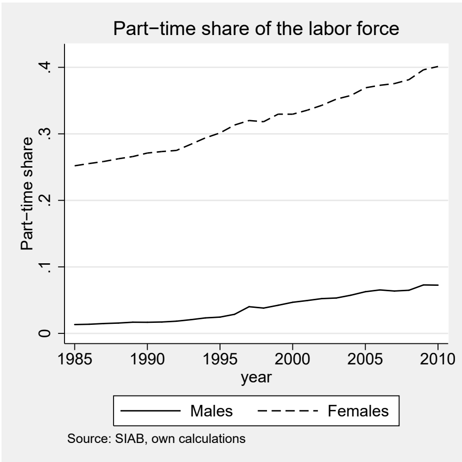
This paper scrutinizes the contribution of composition changes in education, potential work experience, labor market history, industry structure, and occupation on the rise in inequality of full-time wages in Germany from 1985 until 2010. We account explicitly for the growing importance of employment interruptions and temporary part-time episodes among full-time workers, and we estimate the counterfactual full-time wage distribution for all employees.

Our results imply that changes in observables account for a large part of the rise in wage inequality, and that the the growing importance of employment interruptions and temporary part-time episodes play an important role for wage inequality among full-time workers. For males we find that (depending on the base year) 43 to 53 percent of the rise in wage inequality between 1985 and 2010 can be explained by compositional effects of the observables considered. For females the importance of composition changes is even higher, ranging between 64 and 78 percent. To the best of our knowledge the literature has so far not recognized the strong role of composition effects for the rise of female wage inequality. For males composition changes in education (especially in the upper part of the distribution) and changes in recent labor market histories (especially in the lower part of the distribution) are the main contributors to compositional change. The compositional effects of male labor market histories to rising overall wage inequality range from 14 to 17 percent, and from 18 to 23 percent in the lower half of the distribution. For females we find strong composition effects of changes in age/experience and in recent labor market histories. The latter contribute 17 to 18 percent to the overall increase in female wage inequality over the period 1985 to 2010. When including part-timers the role of recent labor market histories becomes even stronger. Our results are policy relevant because both changes in the age/education structure and in labor market histories are observable and to a certain extent predictable. One might wonder to what extent the contribution of increasing heterogeneity in recent labor market histories is causal or to what extent these are just proxies for unobservables. Still, while we are not in a position to separate between these two explanations, accounting for labor market history in fact also proxies for remaining unobservable differences in employment outcomes. Furthermore, the observed trends in previous part-time work and employment interruptions are very strong, which suggests that observed changes in labor market history are mostly the intended consequences of policy changes (section 2.3.2). It is well documented in the literature that part-time work and previous nonemployment have effects on subsequent wages, even when controlling for unobservables (Arulampalam 2001, Schmieder et al. 2010, Paul 2016, Blundell et al. 2016). We therefore expect trends in these variables to directly change the wage distribution in subsequent periods. We also note that our base and target years (1985 and 2010) represent similar points in the business-cycle so that our analysis is unlikely to be affected by huge differences with respect to this aspect. Finally, we note that even if the observed changes in previous part-time work and nonemployment involve increased sorting in terms of unobservables across individuals with differing labor market histories, this would still make histories very relevant factors as their direct effect would be enhanced by changes in unobservables.

2.6 Appendix

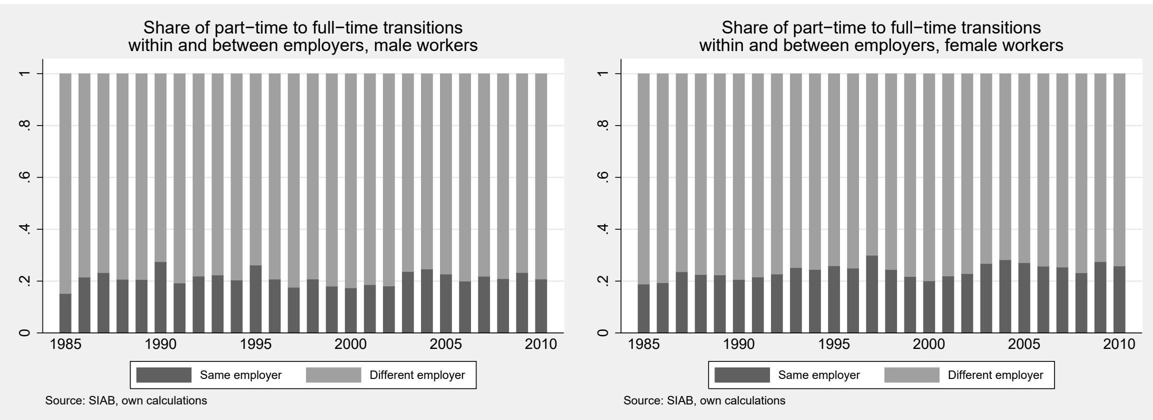
2.6.1 Figures

Figure A2.1: Part-time share



Note: Between 1985 and 2010, the share of the labor force in part-time employment has risen from 26% to 40.3% among women and from 1.9% to 7.8% among men.

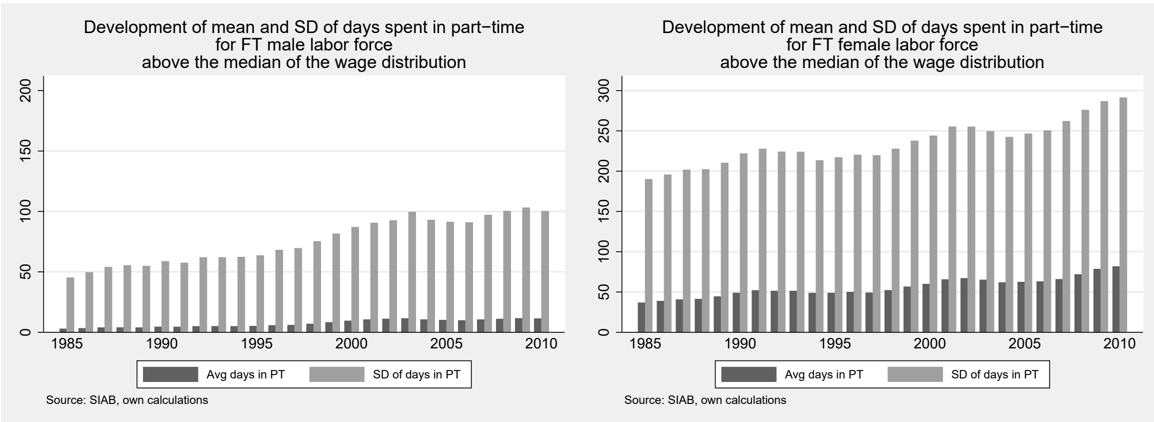
Figure A2.2: Part-time to full-time transitions



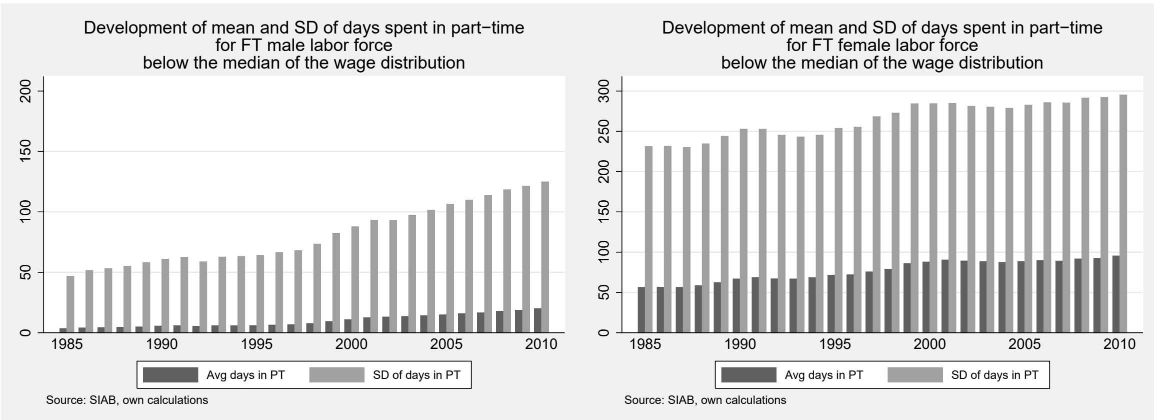
Note: The vast majority of transitions between part-time and full-time jobs involve a change in employers.

Figure A2.3: Days spent in part-time work during the last 5 years (full-timers aged 25-60 years)

Above the median



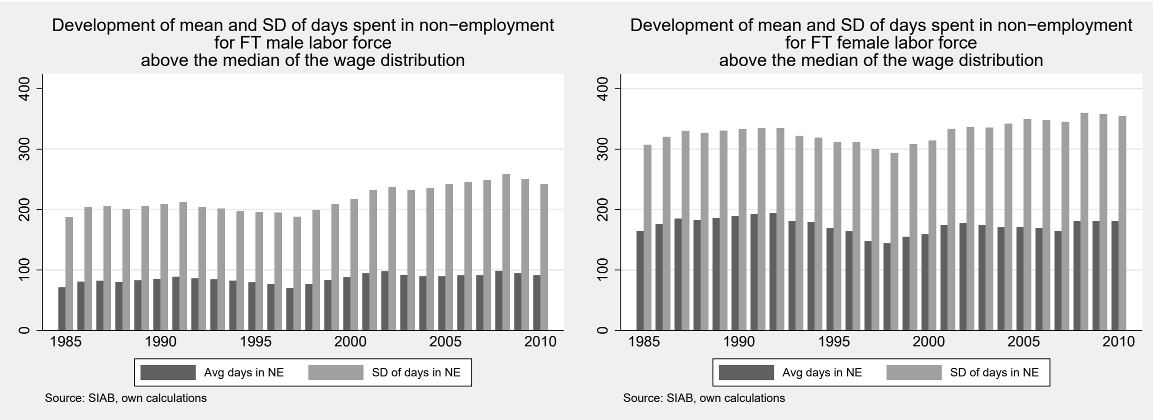
Below the median



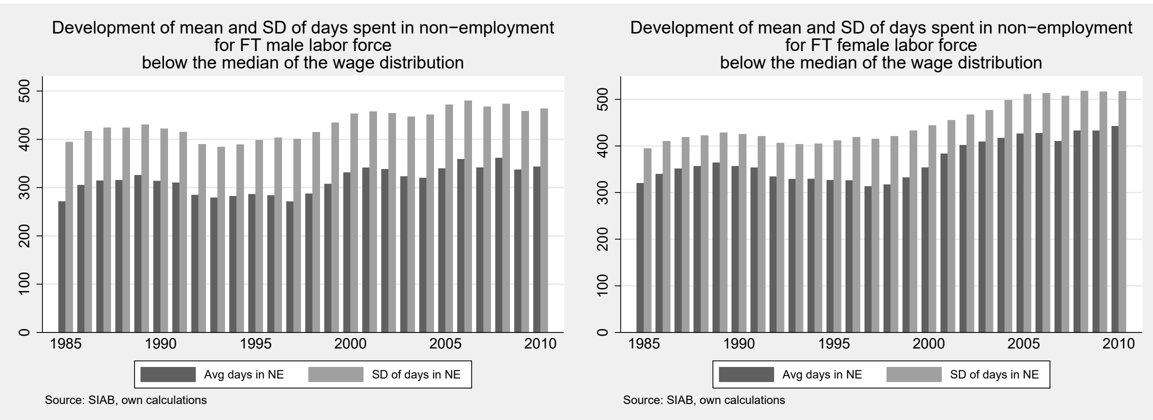
Note: Among current full-time employees between the ages of 25-60, the average number of days historically spent in part-time has increased between 1985 and 2010, particularly among workers earning wages below the median. At the same time, the variance of transient part-time episodes has increased, which implies rising heterogeneity with respect to historic part-time experience.

Figure A2.4: Days spent in nonemployment during the last 5 years (full-timers aged 25-60 years)

Above the median

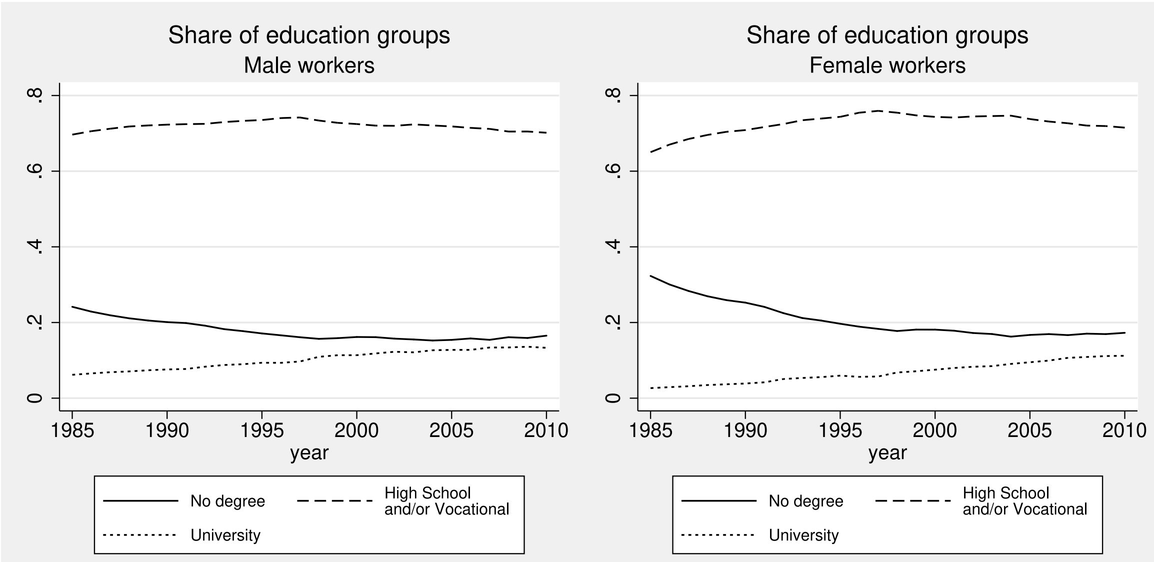


Below the median



Note: Among current full-time employees between the ages of 25-60, the average number of days historically spent in non-employment has increased slightly between 1985 and 2010, particularly among female workers earning wages below the median.

Figure A2.5: Share of education groups

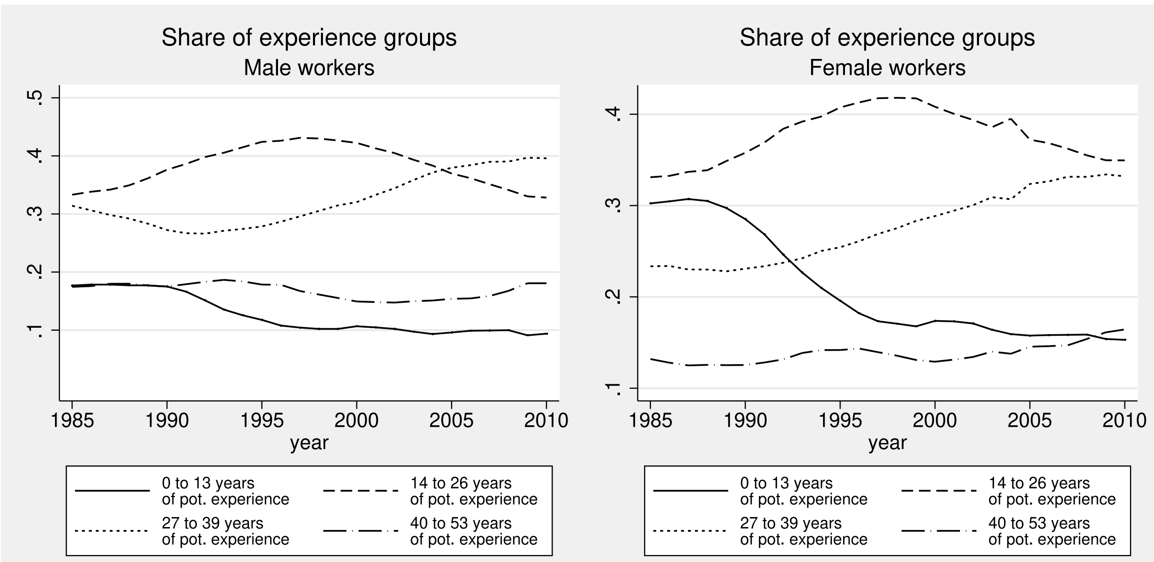


Source: SIAB, own calculations

Source: SIAB, own calculations

Note: Educational expansion has led to consistently increasing shares of university educated workers in the labor force. Simultaneously, the share of workers with low or no qualification has decreased dramatically, especially among women.

Figure A2.6: Share of experience groups

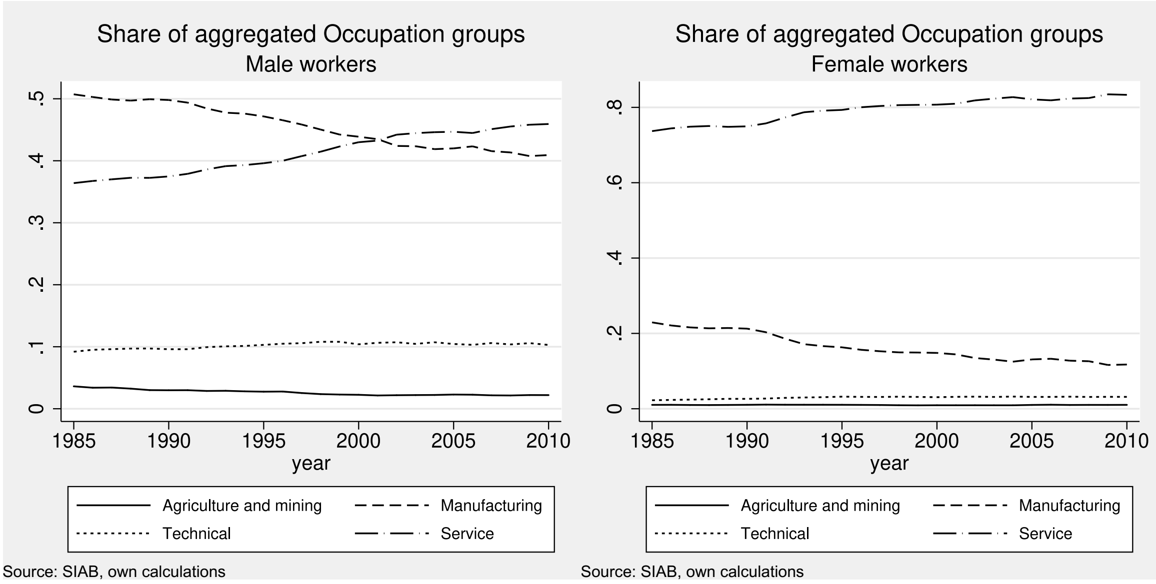


Source: SIAB, own calculations

Source: SIAB, own calculations

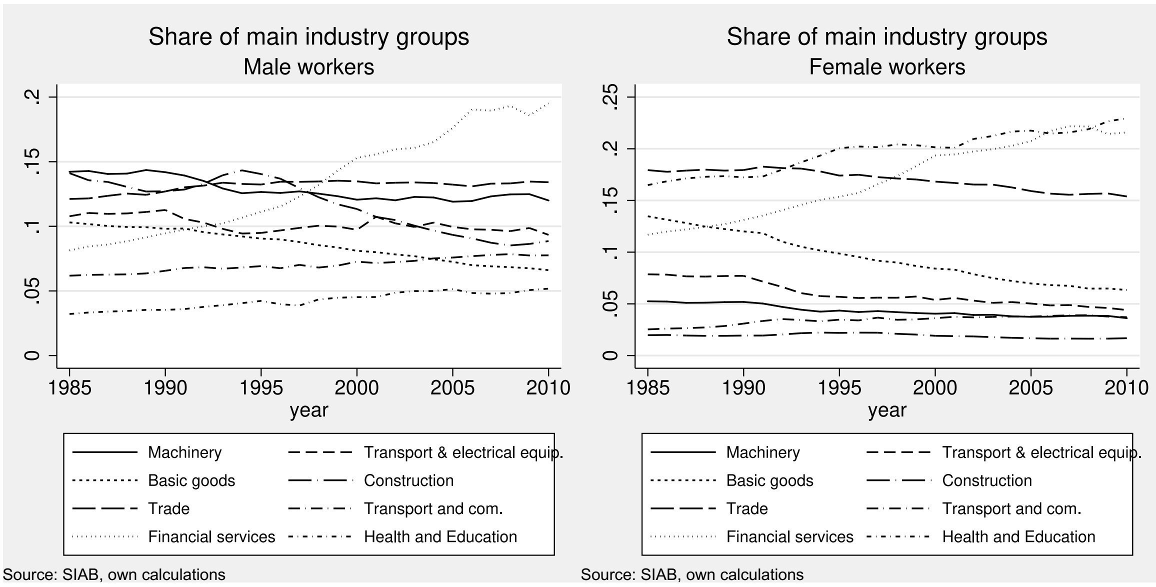
Note: An aging of the labor force results in increasing shares of workers with long time spans to potentially amass work experience. Note particularly the drastic decline in female workers fresh out of education.

Figure A2.7: Share of occupation categories



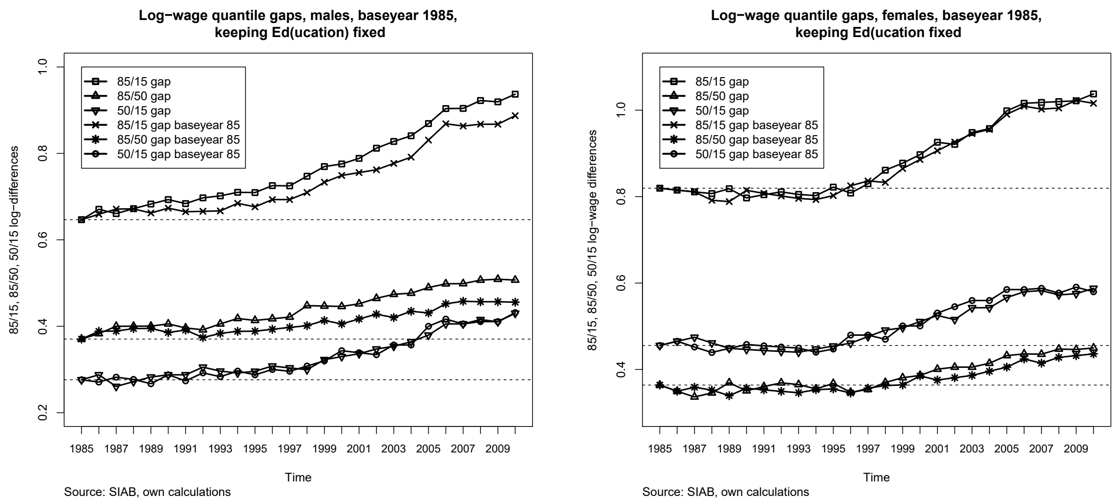
Note: Among both men and women, fewer workers now work in the extended manufacturing industry, even though it is still the second largest employer for male workers. Note particularly that the extended service industry employs the majority of women.

Figure A2.8: Share of industry sectors



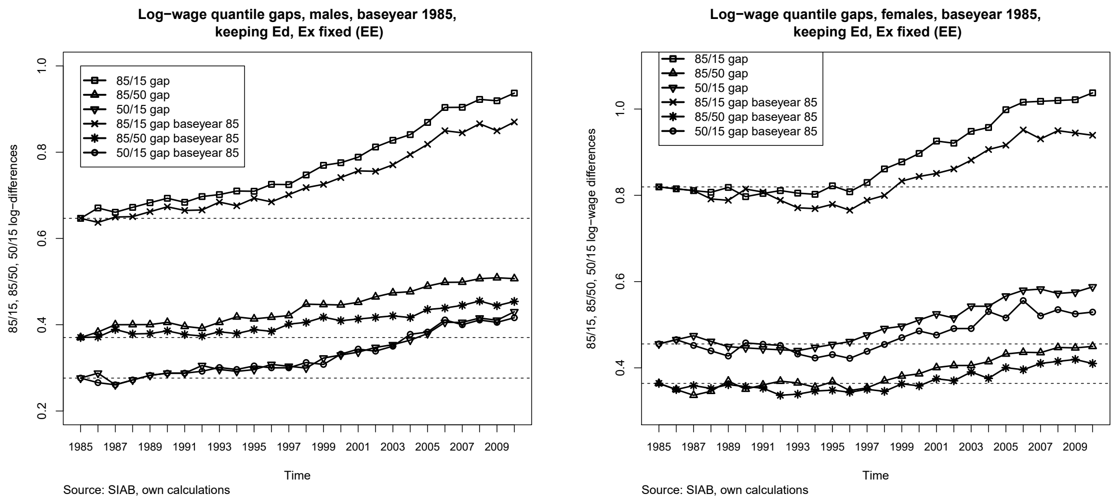
Note: Among both men and women, more workers now work in transport, commerce, financial services, as well as health and education. Simultaneously, fewer workers now work in the basic goods sector and in construction. The mechanical manufacturing sector has shrunk slightly in terms of employment, but not dramatically so.

Figure A2.9: Inequality development base year 1985, specification E (Education)



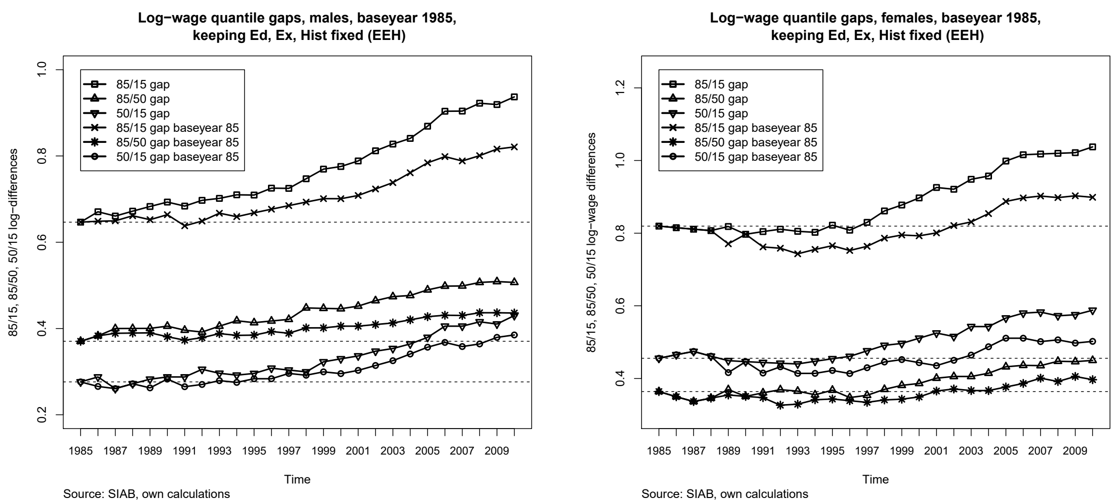
Note: This figure contrasts the observed increase in the gaps between wage quantiles with the counterfactual increase in wage gaps, which would have prevailed if the distribution of education in the workforce had not changed since 1985.

Figure A2.10: Inequality development base year 1985, specification EE (Education, Experience)



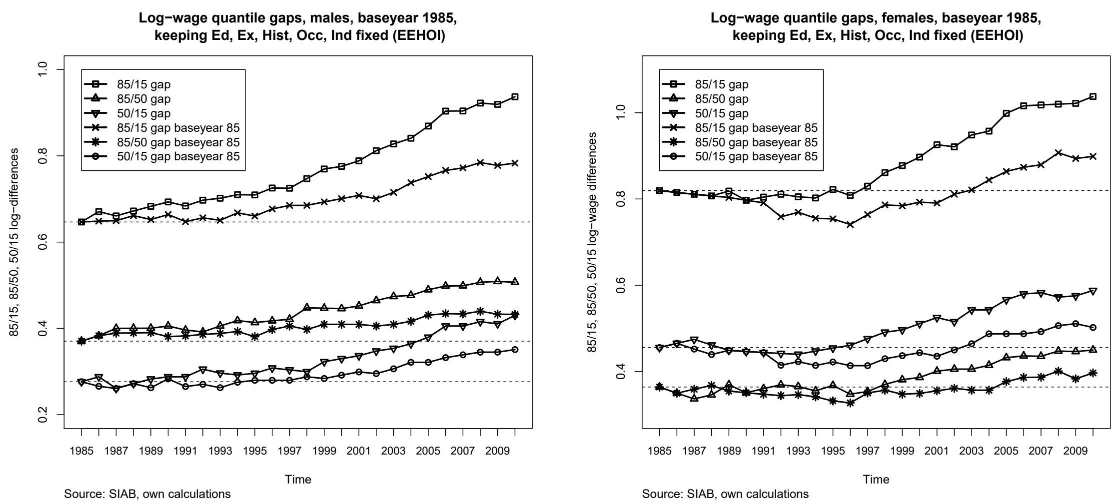
Note: This figure contrasts the observed increase in the gaps between wage quantiles with the counterfactual increase in wage gaps, which would have prevailed if the distribution of education and potential experience in the workforce had not changed since 1985.

Figure A2.11: Inequality development base year 1985, specification EEH (Education, Experience, Labor market history)



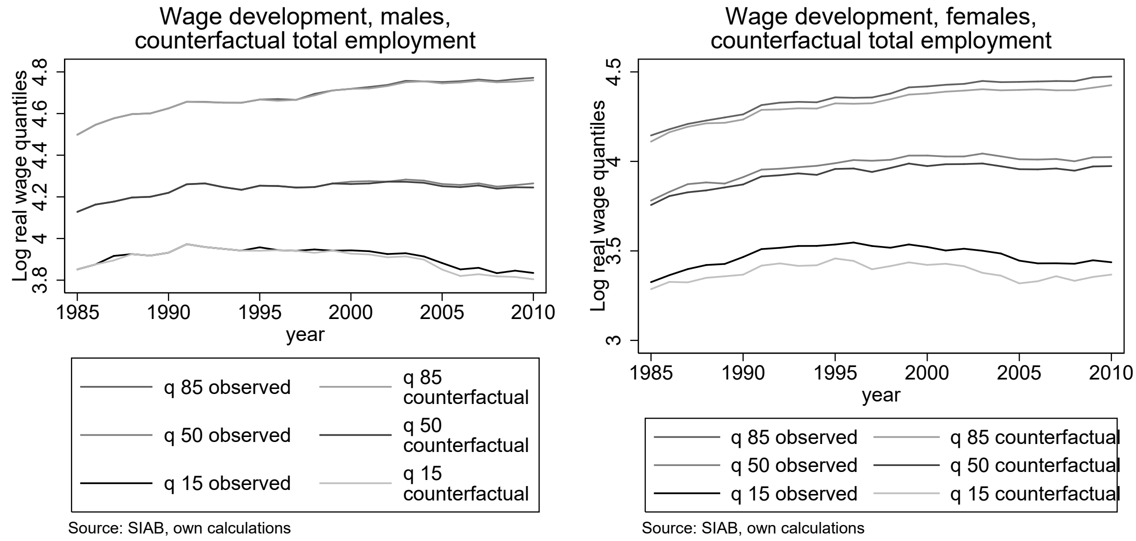
Note: This figure contrasts the observed increase in the gaps between wage quantiles with the counterfactual increase in wage gaps, which would have prevailed if the distribution of education, potential experience and labor market histories in the workforce had not changed since 1985.

Figure A2.12: Inequality development base year 1985, specification EEHOI (Education, Experience, Labor market history, Occupation, Industry sector)



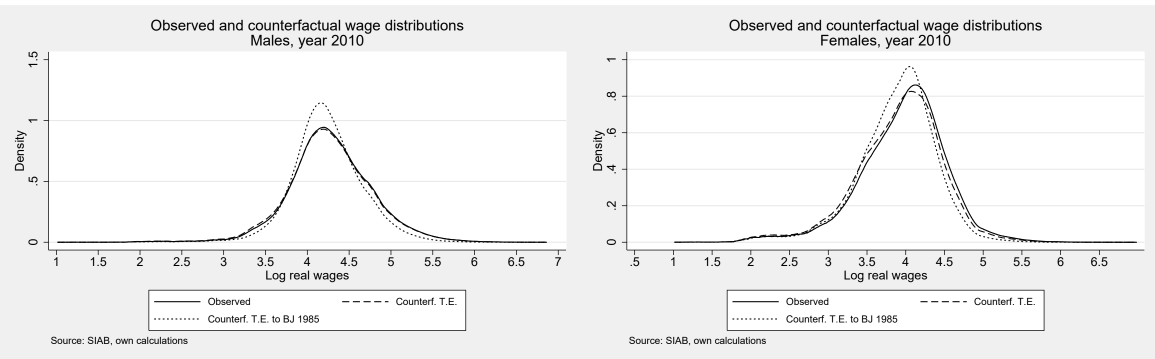
Note: This figure contrasts the observed increase in the gaps between wage quantiles with the counterfactual increase in wage gaps, which would have prevailed if the distribution of education, potential experience, labor market histories, occupations and industry in the workforce had not changed since 1985.

Figure A2.13: Counterfactual wage distribution, if full-timers had total employment characteristics



Note: This figure contrasts the observed log wage quantiles with the counterfactual log wage quantiles if the employed had the characteristics distribution of the entire labor force (employed and unemployed)

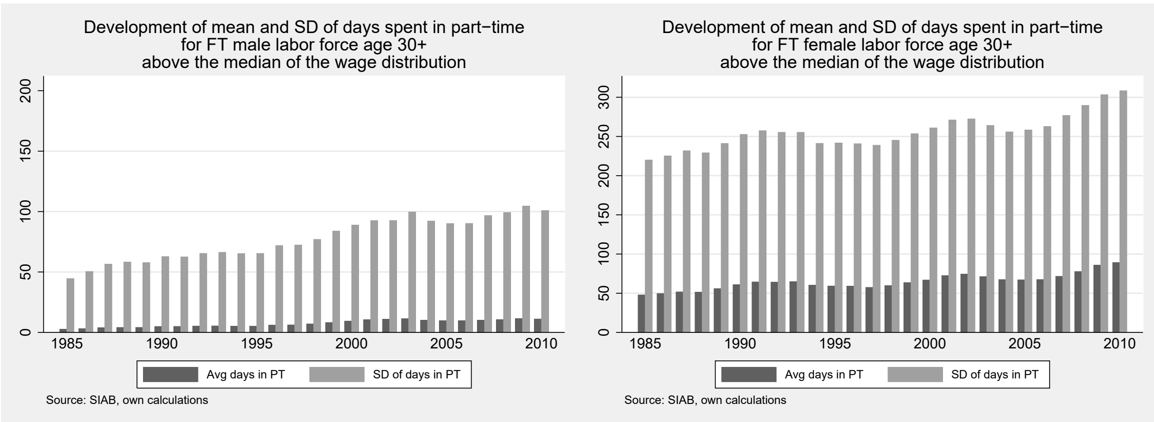
Figure A2.14: Comparison of observed, counterfactual total employment and reweighted counterfactual total employment sample (specification EEHOI)



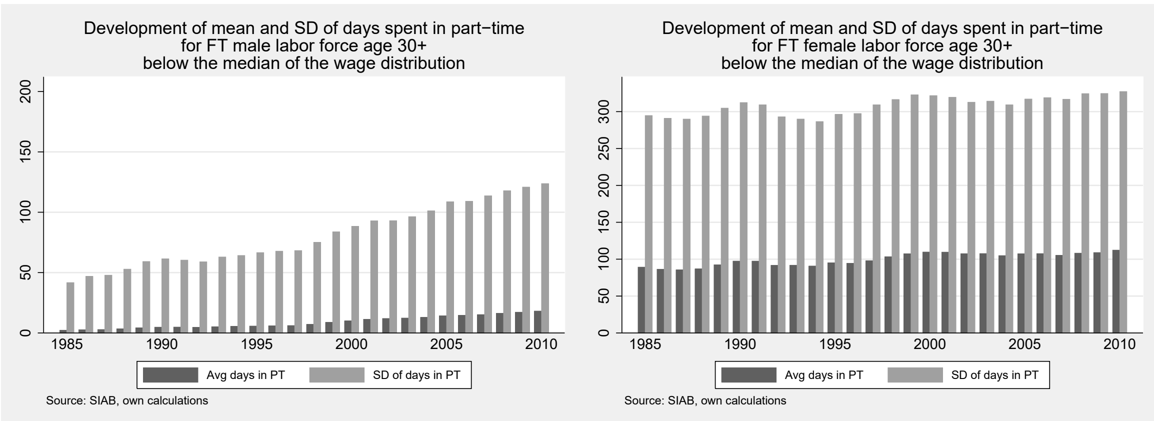
Note: The wage distribution under counterfactual total employment in 2010 would prevail if the employed had the characteristics distribution of the entire labor force in 2010. Counterfactual total employment of 1985 reflects the wage distribution which would have prevailed if workers in 2010 had the characteristics distribution of the entire labor force of 1985 (employed and unemployed).

Figure A2.15: Days spent in part-time work during the last 5 years (full-timers aged 30-60 years)

Above the median



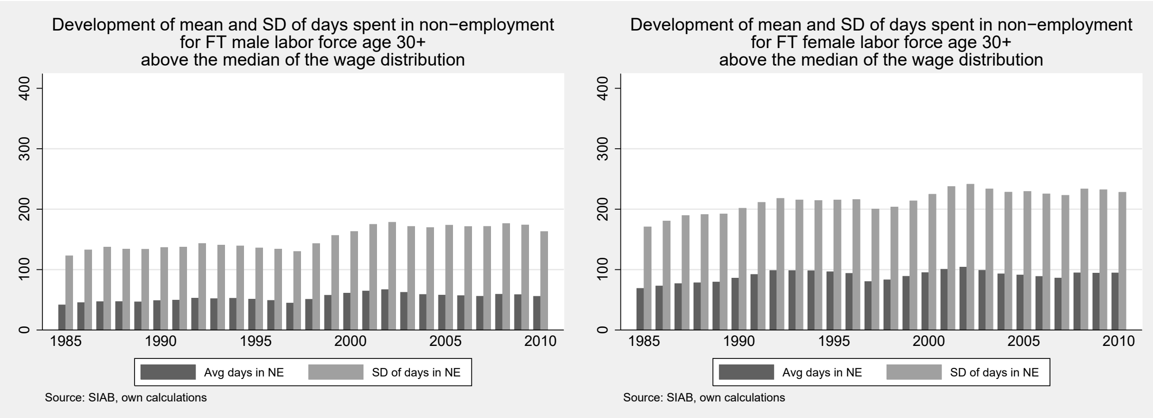
Below the median



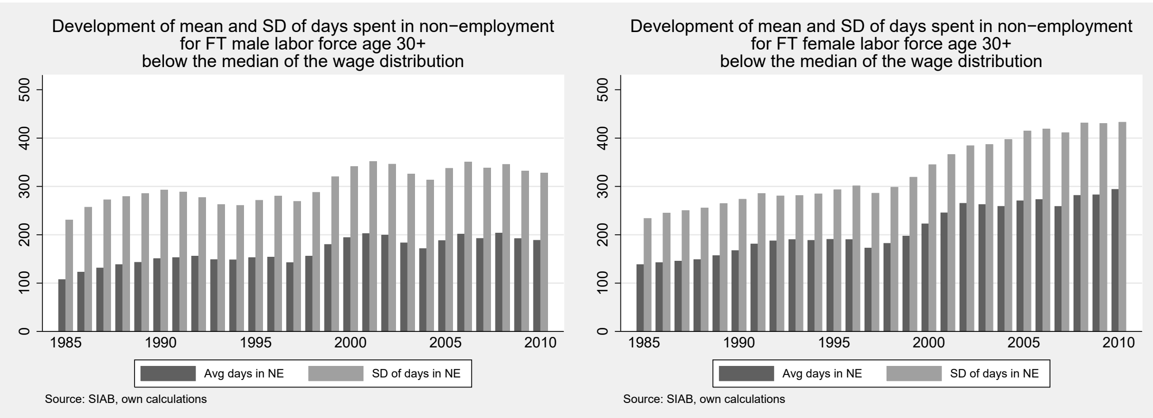
Note: Among current full-time employees between the ages of 30-60, the average number of days historically spent in part-time has increased between 1985 and 2010. At the same time, the variance of transient part-time episodes has increased, which implies rising heterogeneity with respect to historic part-time experience.

Figure A2.16: Days spent in nonemployment during the last 5 years (full-timers aged 30-60 years)

Above the median

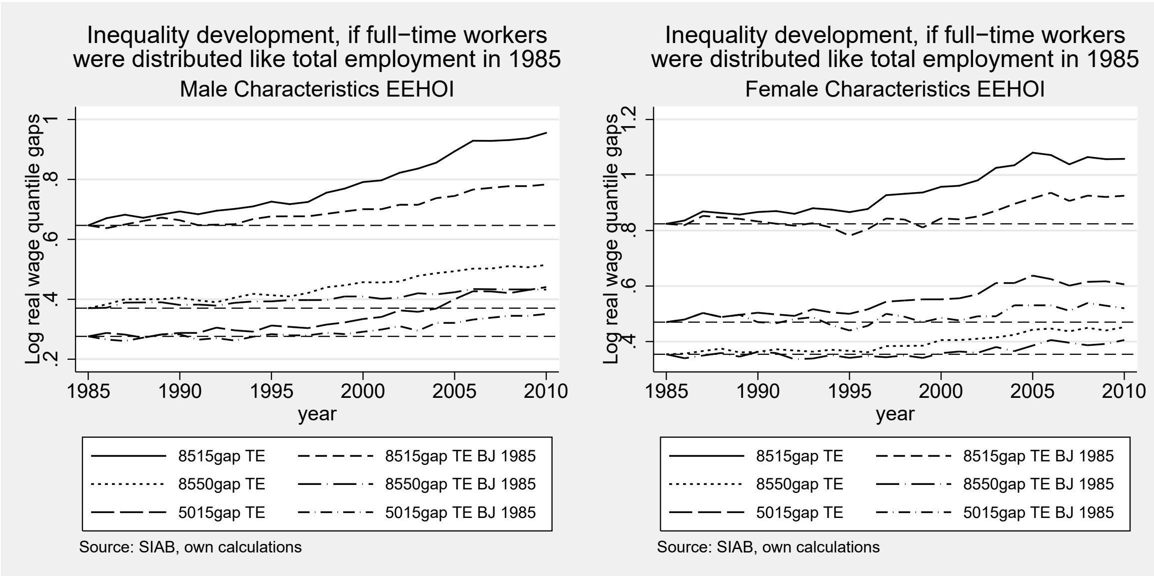


Below the median



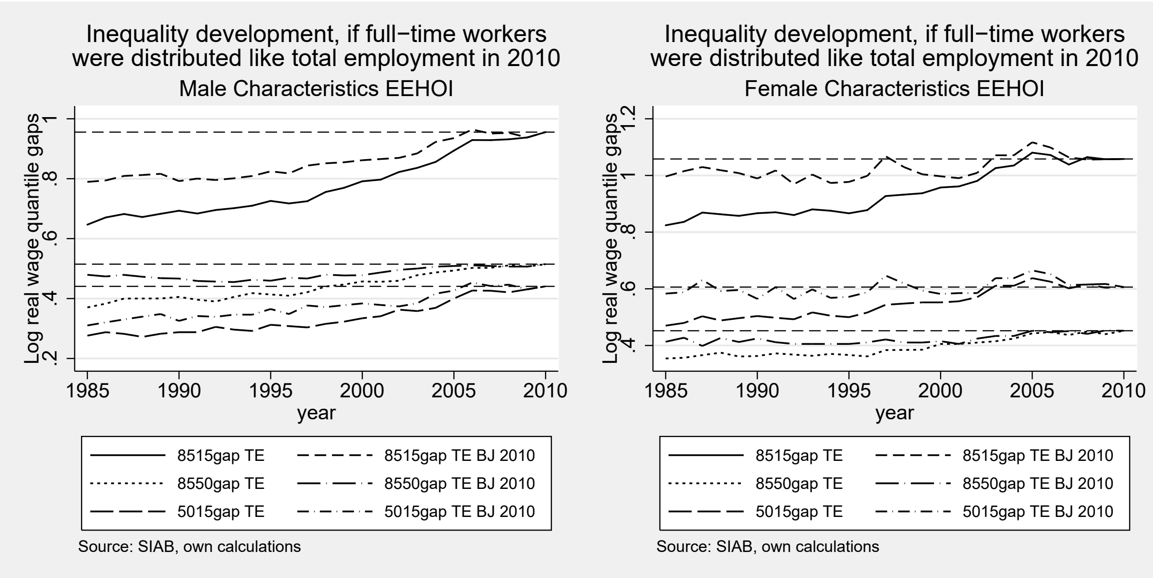
Note: Among current full-time employees between the ages of 30-60, the average number of days historically spent outside of employment, and their variance, have increased between 1985 and 2010.

Figure A2.17: Inequality development base year 1985, specification EEHOI of total employment



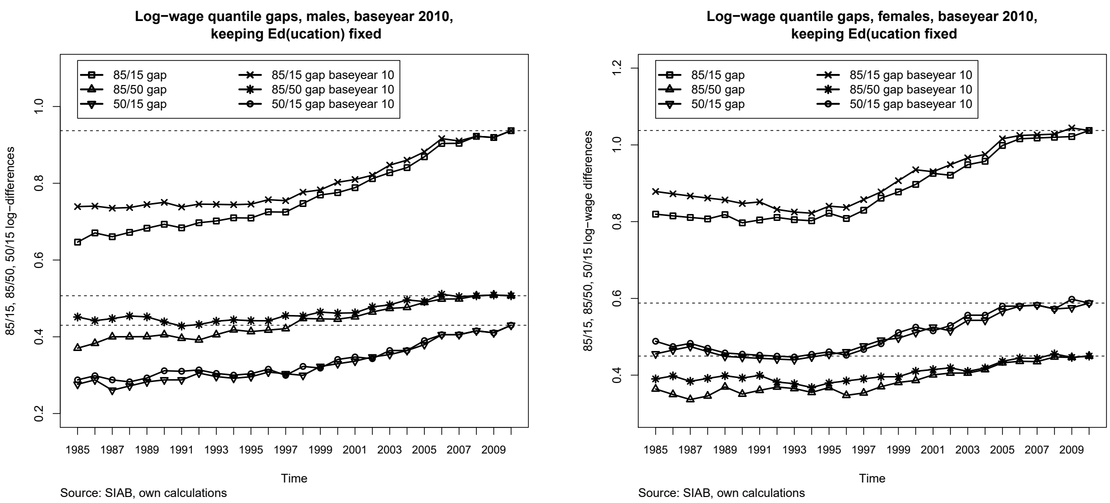
Note: This figure contrasts the counterfactual wage gaps which would have prevailed if the employed had the characteristics of the entire labor force in the respective year (employed and unemployed) with the counterfactual wage gaps which would have prevailed if the employed had the characteristics of the entire labor force of 1985.

Figure A2.18: Inequality development base year 2010, specification EEHOI of total employment



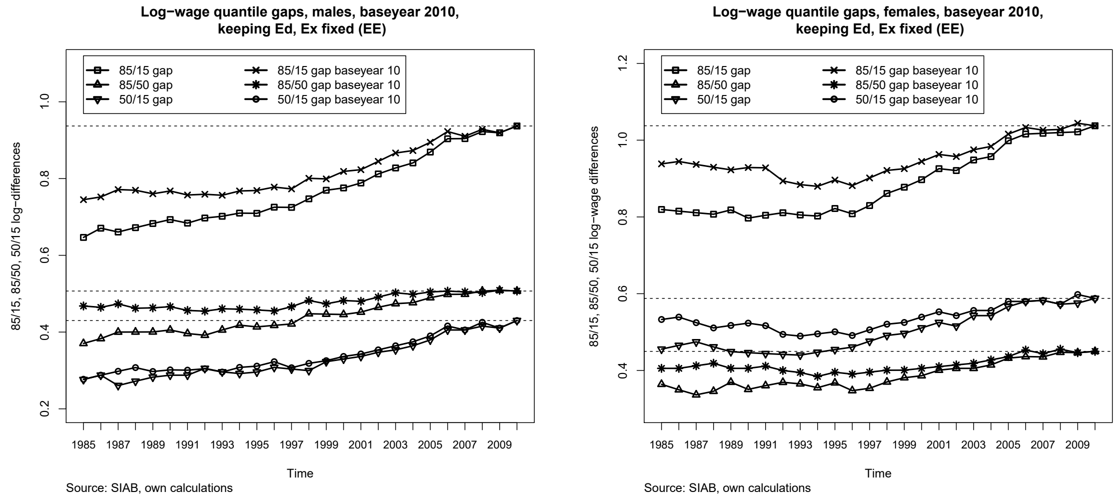
Note: This figure contrasts the counterfactual wage gaps which would have prevailed if the employed had the characteristics of the entire labor force in the respective year (employed and unemployed) with the counterfactual wage gaps which would have prevailed if the employed had the characteristics of the entire labor force of 2010.

Figure A2.19: Inequality development base year 2010, specification E



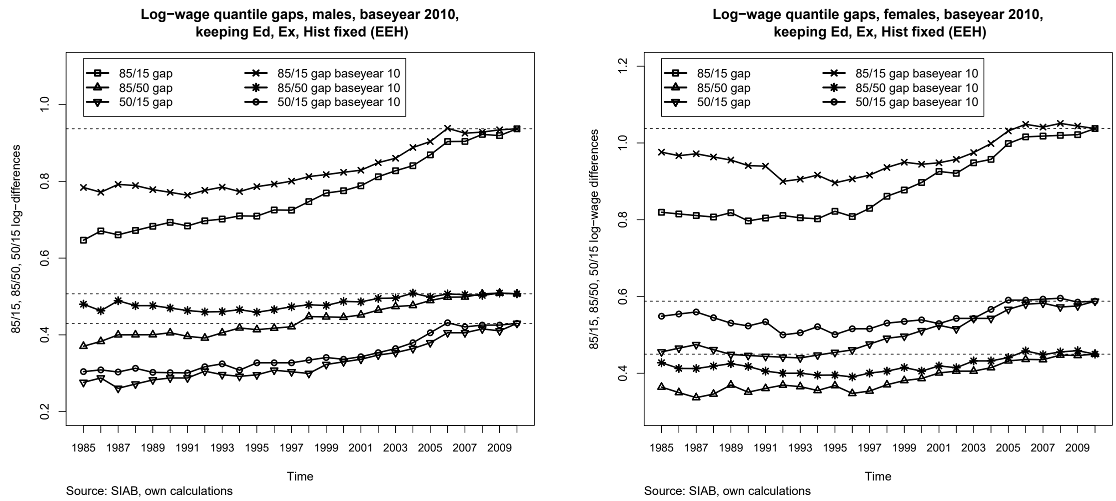
Note: This figure contrasts the observed increase in the gaps between wage quantiles with the counterfactual increase in wage gaps which would have prevailed if the distribution of education in the workforce always been that of 2010.

Figure A2.20: Inequality development base year 2010, specification EE



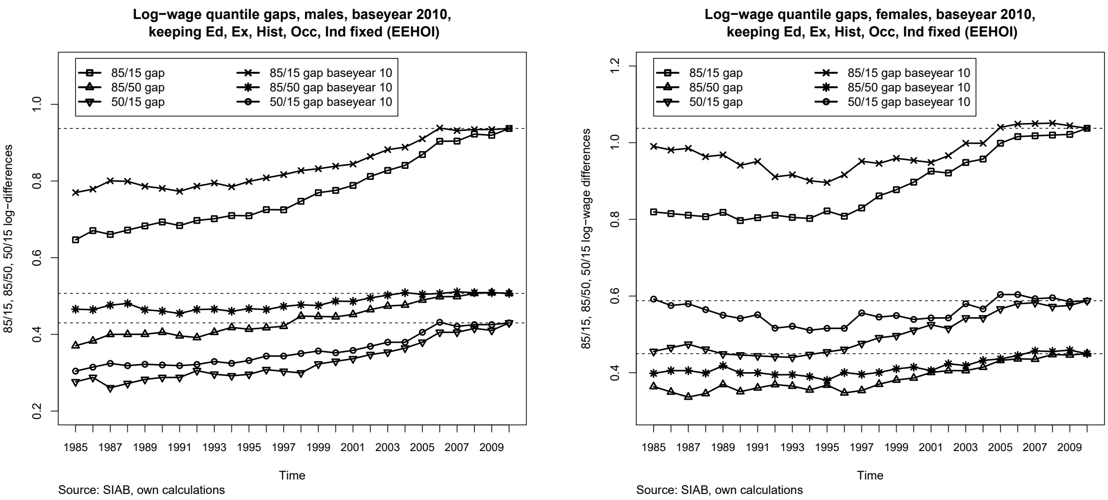
Note: This figure contrasts the observed increase in the gaps between wage quantiles with the counterfactual increase in wage gaps which would have prevailed if the distribution of education and potential experience in the workforce always been that of 2010.

Figure A2.21: Inequality development base year 2010, specification EEH



Note: This figure contrasts the observed increase in the gaps between wage quantiles with the counterfactual increase in wage gaps which would have prevailed if the distribution of education, potential experience and labor market histories in the workforce always been that of 2010.

Figure A2.22: Inequality development base year 2010, specification EEHOI



Note: This figure contrasts the observed increase in the gaps between wage quantiles with the counterfactual increase in wage gaps which would have prevailed if the distribution of education, potential experience, labor market histories, occupations and industry in the workforce always been that of 2010.

2.6.2 Tables

Table A2.1: Descriptives of full-time samples

Male full-time sample

	1985		2010	
	mean	sd	mean	sd
Real wage in Euro	70.06	47.53	82.48	48.34
Log real wage	4.16	0.39	4.28	0.51
No/other degree indicator	0.19	0.40	0.08	0.28
Vocational degree indicator	0.71	0.46	0.71	0.45
University degree indicator	0.07	0.25	0.15	0.36
Work experience	27.34	11.19	28.98	10.13
No. of days in full time last 5 years	1546.04	487.51	1523.88	513.84
Fulltime spell in previous year?	0.96	0.19	0.96	0.20
No. of days in part time last 5 years	3.26	46.49	15.72	113.47
Part-time spell in previous year?	0.00	0.05	0.01	0.09
Agriculture and mining	0.03	0.17	0.02	0.13
Plastics, rubber, mineral products	0.03	0.17	0.03	0.17
Chemicals	0.03	0.18	0.02	0.15
Machinery and metal products	0.15	0.36	0.13	0.33
Transport- and electrical equipment	0.12	0.32	0.11	0.31
Food and basic consumption	0.10	0.31	0.07	0.25
Hotels and restaurants	0.01	0.11	0.02	0.13
Construction	0.12	0.32	0.08	0.28
Trade	0.12	0.33	0.14	0.35
Transport and communication	0.06	0.24	0.07	0.26
Financial and insurance	0.08	0.27	0.18	0.38
Public services	0.04	0.20	0.05	0.21
Health and Education	0.03	0.18	0.05	0.23
Public administration	0.06	0.24	0.04	0.20

Female full-time sample

	1985		2010	
	mean	sd	mean	sd
Real wage in Euro	46.24	20.34	61.02	35.16
Log real wage	3.74	0.44	3.97	0.56
No/other degree indicator	0.27	0.45	0.08	0.27
Vocational degree indicator	0.66	0.47	0.73	0.44
University degree indicator	0.03	0.16	0.12	0.32
Work experience	24.03	11.90	27.39	11.14
No. of days in full time last 5 years	1356.01	598.16	1327.35	625.94
Fulltime spell in previous year?	0.93	0.26	0.93	0.26
No. of days in part time last 5 years	45.97	210.01	88.99	292.80
Part-time spell in previous year?	0.02	0.15	0.04	0.20
Agriculture and mining	0.01	0.08	0.01	0.08
Plastics, rubber, mineral products	0.02	0.14	0.01	0.11
Chemicals	0.02	0.15	0.02	0.13
Machinery and metal products	0.05	0.23	0.04	0.19
Transport- and electrical equipment	0.08	0.27	0.05	0.21
Food and basic consumption	0.13	0.34	0.06	0.25
Hotels and restaurants	0.03	0.18	0.03	0.18
Construction	0.02	0.14	0.02	0.13
Trade	0.18	0.38	0.16	0.36
Transport and communication	0.02	0.16	0.04	0.19
Financial and insurance	0.12	0.33	0.21	0.40
Public services	0.06	0.23	0.06	0.24
Health and Education	0.17	0.38	0.24	0.42
Public administration	0.08	0.27	0.06	0.24

Table A2.2: Male's most Frequent Occupations 1985 and 2010 (Top 5)

	1985		2010	
	Occupation	Share	Occupation	Share
Rank 1	Transportation	5.8%	Office workers	7.8%
Rank 2	Metalworkers	5.8%	Transportation	6%
Rank 3	Office workers	5.6%	Storage workers	5.1%
Rank 4	Technicians	5.5%	Retail workers	5.1%
Rank 5	Storage workers	4.8%	Technicians	5.0%

Table A2.3: Female's most Frequent Occupations 1985 and 2010 (Top 5)

	1985		2010	
	Occupation	Share	Occupation	Share
Rank 1	Office workers	25.5%	Office workers	26.9%
Rank 2	Retail workers	11.3%	Healthcare	12.4%
Rank 3	Healthcare	9.2%	Retail workers	9.8%
Rank 4	Assembly workers	4.1%	Social workers	6.9%
Rank 5	Cleaning	3.7%	Banking & Insurance	3.5%

Table A2.4: Reweighted inequality increase, 1985-2010, males, compositions for base year 1985

	Observed	Ed			Ed+Ex			Ed+Ex+Hist			Ed+Ex+Hist+Occ+Ind		
	Increase	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment
85/15	0.290	0.241	17.11%	17.11%	0.223	23.04%	5.93%	0.174	39.95%	16.92%	0.137	52.97%	13.02%
85/50	0.137	0.085	37.50%	37.50%	0.084	38.82%	1.32%	0.066	52.00%	13.18%	0.062	54.64%	2.64%
50/15	0.154	0.155	-1.00%	-1.00%	0.140	9.02%	10.02%	0.109	29.25%	19.23%	0.075	51.49%	22.24%
90/10 residual	0.183	0.170	7.07%	7.07%	0.165	9.71%	2.64%	0.146	20.37%	10.66%	0.121	33.96%	13.59%

Table A2.5: Reweighted inequality increase, 1985-2010, females, compositions for base year 1985

	Observed	Ed			Ed+Ex			Ed+Ex+Hist			Ed+Ex+Hist+Occ+Ind		
	Increase	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment
85/15	0.218	0.196	9.91%	9.91%	0.120	45.05%	35.14%	0.079	63.64%	18.59%	0.079	63.64%	0.00%
85/50	0.086	0.072	15.90%	15.90%	0.046	46.30%	30.40%	0.033	61.89%	15.59%	0.033	61.89%	0.00%
50/15	0.132	0.124	6.02%	6.02%	0.074	44.23%	38.21%	0.047	64.77%	20.54%	0.047	64.77%	0.00%
90/10 residual	0.185	0.193	-4.38%	-4.38%	0.132	28.77%	33.15%	0.100	45.89%	17.12%	0.090	51.56%	5.67%

Table A2.6: Change in inequality measures since 1985, for males, composition adjusted to total employment in base year 1985

	Observed	Ed			Ed+Ex			Ed+Ex+Hist			Ed+Ex+Hist+Occ+Ind		
	Increase	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment
85/15	0.309	0.257	16.90%	16.90%	0.249	19.24%	2.34%	0.174	43.55%	24.31%	0.137	55.79%	12.24%
85/50	0.144	0.096	33.72%	33.72%	0.087	39.43%	5.71%	0.076	47.43%	8.00%	0.062	57.11%	9.68%
50/15	0.164	0.161	2.14%	2.14%	0.162	1.51%	-0.63%	0.098	40.15%	38.64%	0.075	54.64%	14.49%
90/10 residual	0.195	0.183	6.23%	6.23%	0.177	8.92%	2.69%	0.143	26.75%	17.83%	0.117	39.72%	12.97%

Table A2.7: Change in inequality measures since 1985, for females, composition adjusted to total employment in base year 1985

	Observed	Ed			Ed+Ex			Ed+Ex+Hist			Ed+Ex+Hist+Occ+Ind		
	Increase	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment
85/15	0.234	0.222	5.19%	5.19%	0.150	35.72%	30.53%	0.101	56.76%	21.04%	0.101	56.76%	0.00%
85/50	0.098	0.093	5.25%	5.25%	0.075	23.16%	17.91%	0.065	33.65%	10.49%	0.051	47.56%	13.91%
50/15	0.136	0.129	5.15%	5.15%	0.075	44.74%	39.59%	0.036	73.36%	28.62%	0.050	63.37%	-9.99%
90/10 residual	0.158	0.173	-9.09%	-9.09%	0.114	28.31%	37.40%	0.078	50.44%	22.31%	0.069	56.12%	5.68%

Table A2.8: Tests of differences between labor market histories above and below the median wage

		Days in part-time over the last 5 years		Days in nonemployment over the last 5 years	
	Year	1985	2010	1985	2010
Males	P-value of difference in means	0.0035	0	0	0
	P-value of difference in variances	0.0067	0	0	0
Females	P-value of difference in means	0	0	0	0
	P-value of difference in variances	0	0	0	0

Table A2.9: Descriptive statistics of combined full-time and part-time samples

Males

	1985		2010	
	mean	sd	mean	sd
Real wage in Euro	69.89	47.51	81.61	48.15
Log real wage	4.15	0.40	4.27	0.52
No/other degree indicator	0.19	0.40	0.08	0.28
Vocational degree indicator	0.70	0.46	0.71	0.46
University degree indicator	0.07	0.25	0.16	0.36
Work experience	27.32	11.19	29.08	10.23
No. of days in full time last 5 years	1540.45	494.58	1494.81	544.03
Fulltime spell in previous year?	0.96	0.20	0.94	0.24
No. of days in part time last 5 years	6.29	78.41	37.03	203.05
Part-time spell in previous year?	0.01	0.08	0.03	0.18
Agriculture and mining	0.03	0.17	0.02	0.13
Plastics, rubber, mineral products	0.03	0.17	0.03	0.17
Chemicals	0.03	0.18	0.02	0.15
Machinery and metal products	0.15	0.36	0.13	0.33
Transport- and electrical equipment	0.12	0.32	0.10	0.31
Food and basic consumption	0.10	0.31	0.07	0.25
Hotels and restaurants	0.01	0.11	0.02	0.14
Construction	0.12	0.32	0.08	0.27
Trade	0.12	0.33	0.14	0.35
Transport and communication	0.06	0.24	0.07	0.26
Financial and insurance	0.08	0.27	0.18	0.38
Public services	0.04	0.20	0.05	0.21
Health and Education	0.04	0.18	0.06	0.23
Public administration	0.06	0.24	0.04	0.20

Females

	1985		2010	
	mean	sd	mean	sd
Real wage in Euro	43.64	20.41	55.40	33.52
Log real wage	3.67	0.48	3.85	0.59
No/other degree indicator	0.28	0.45	0.08	0.28
Vocational degree indicator	0.65	0.48	0.73	0.44
University degree indicator	0.03	0.16	0.11	0.32
Work experience	24.99	11.90	28.62	10.99
No. of days in full time last 5 years	1199.98	696.74	1048.65	759.52
Fulltime spell in previous year?	0.81	0.39	0.72	0.45
No. of days in part time last 5 years	209.50	513.80	366.32	642.41
Part-time spell in previous year?	0.14	0.35	0.25	0.43
Agriculture and mining	0.01	0.08	0.01	0.08
Plastics, rubber, mineral products	0.02	0.14	0.01	0.10
Chemicals	0.02	0.14	0.01	0.12
Machinery and metal products	0.05	0.22	0.03	0.18
Transport- and electrical equipment	0.07	0.26	0.04	0.20
Food and basic consumption	0.12	0.33	0.06	0.23
Hotels and restaurants	0.03	0.18	0.03	0.18
Construction	0.02	0.14	0.02	0.13
Trade	0.19	0.39	0.16	0.37
Transport and communication	0.03	0.17	0.04	0.19
Financial and insurance	0.12	0.33	0.19	0.40
Public services	0.06	0.23	0.06	0.24
Health and Education	0.18	0.38	0.26	0.44
Public administration	0.08	0.28	0.07	0.26

Table A2.10: Change in inequality measures since 1985, for males, composition adjusted to total employment in base year 2010

	Observed	Ed			Ed+Ex			Ed+Ex+Hist			Ed+Ex+Hist+Occ+Ind		
	Increase	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment
85/15	0.309	0.222	36.71%	36.71%	0.215	39.24%	2.53%	0.159	58.32%	19.08%	0.166	55.96%	-2.36%
85/50	0.144	0.069	60.46%	60.46%	0.051	71.68%	11.22%	0.045	75.67%	3.99%	0.036	81.00%	5.33%
50/15	0.164	0.153	10.64%	10.64%	0.164	0.71%	-9.93%	0.114	41.36%	40.65%	0.131	29.29%	-12.07%
90/10 residual	0.195	0.160	-4.04%	-4.04%	0.149	4.87%	8.91%	0.056	68.78%	63.91%	0.058	67.67%	-1.11%

Table A2.11: Change in inequality measures since 1985, for females, composition adjusted to total employment in base year 2010

	Observed	Ed			Ed+Ex			Ed+Ex+Hist			Ed+Ex+Hist+Occ+Ind		
	Increase	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment
85/15	0.234	0.173	31.10%	31.10%	0.087	68.30%	37.20%	0.035	88.04%	19.74%	0.062	78.15%	-9.89%
85/50	0.098	0.047	58.46%	58.46%	0.047	58.46%	0.00%	0.031	73.55%	15.09%	0.039	66.00%	-7.55%
50/15	0.136	0.127	8.88%	8.88%	0.041	75.07%	66.19%	0.004	97.73%	22.66%	0.023	86.39%	-11.34%
90/10 residual	0.158	0.156	-15.97%	-15.97%	0.077	47.55%	63.52%	0.036	76.60%	29.05%	0.018	88.34%	11.74%

Table A2.12: Reweighted inequality increase 1985-2010, males, compositions for base year 2010

	Observed	Ed			Ed+Ex			Ed+Ex+Hist			Ed+Ex+Hist+Occ+Ind		
	Increase	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment
85/15	0.290	0.198	31.87%	31.87%	0.192	33.87%	2.00%	0.153	47.34%	13.47%	0.167	42.46%	-4.88%
85/50	0.137	0.055	59.40%	59.40%	0.039	71.47%	12.07%	0.027	80.17%	8.70%	0.041	69.80%	-10.37%
50/15	0.154	0.142	7.43%	7.43%	0.153	0.48%	-6.95%	0.126	18.18%	17.70%	0.126	18.18%	0.00%
90/10 residual	0.183	0.149	18.64%	18.64%	0.138	24.62%	5.98%	0.107	41.36%	16.74%	0.094	48.73%	7.37%

Table A2.13: Reweighted inequality increase 1985-2010, females, compositions for base year 2010

	Observed	Ed			Ed+Ex			Ed+Ex+Hist			Ed+Ex+Hist+Occ+Ind		
	Increase	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment	Remaining increase	Explained share	Increment
85/15	0.218	0.159	27.10%	27.10%	0.099	54.49%	27.39%	0.062	71.79%	17.30%	0.047	78.39%	6.60%
85/50	0.086	0.060	30.58%	30.58%	0.044	48.38%	17.80%	0.022	74.01%	25.63%	0.051	40.09%	-33.92%
50/15	0.132	0.099	24.85%	24.85%	0.055	58.44%	33.59%	0.039	70.36%	11.92%	-0.004	103.22%	32.86%
90/10 residual	0.185	0.180	2.55%	2.55%	0.093	49.56%	47.01%	0.061	66.94%	17.38%	0.046	75.11%	8.17%

2.6.3 Imputation of wages above censoring threshold

Our imputation procedure for wages above the contribution threshold of social security is loosely based on Gartner (2005). We assume that log-wages are approximately normally distributed and estimate expected wages above the censoring point with a Tobit model. We regress log wages on education, age, nationality and individual labor market history, separately for both genders. Results in the literature suggest that this type of imputation leads to a slight upward bias in the variance of wages each year. Important for our analysis, however, it does not lead to bias in the trend of wage dispersion.[15] As we want to take into account that the variance of wages is potentially correlated with individual characteristics, we modify the procedure suggested by Gartner (2005) to explicitly model a heteroscedastic variance for the Tobit regression. A simple imputation of log wages from the Tobit model would exhibit too little variation. We therefore adjust imputed wages by a random draw from a truncated normal distribution, using the predicted heteroscedastic variance from the Tobit model. We impute separately for each year and for male and female workers. Imputation by this method raises the mean wage by 0.8% and the standard deviation 14.6% for males, and by 0.2% as well as 3.2% for females across all years.

2.6.4 Details of the counterfactual analysis

Composition reweighting for full-timers

We account for the selection into full-time work based on the observed composition of workers regarding their socio-economic characteristics. Changes in the composition over time reflect selective movements of individuals into and out of full-time work. Our aim is to quantify the effects of such changes in the composition of full-timers on wage inequality. We use the reweighting methodology introduced by DiNardo et al. (1996) to estimate counterfactual wage distributions fixing the composition of a reference group (in our case the population of full-timers in a reference year).

In the first part of our analysis, we analyze the distribution of full-time wages which would result if the distribution of worker characteristics had not changed over time but only the conditional wage structure (i.e. the wage distribution holding characteristics constant). Based on these counterfactual wage distributions, we calculate and compare the development of inequality as measured by the gaps between the 85th, 50th and the 15th wage percentiles and the spread of residual wages. We take the residuals from a Mincer regression of log

[15] Compare the discussion in Card et al. (2013).

wages w on a flexible specification of the characteristics listed in table 2.1. The dispersion of residual wages represents wage inequality within narrow groups of workers defined by the characteristics given in table 2.1. Changes in residual wage inequality may also be the result of changes in the composition of the labor force (Lemieux, 2006). This will be the case if there is heteroscedasticity, i.e. the conditional residual variance depends on observed characteristics. In this case, shifts in the distribution of characteristics affect residual wage inequality. For instance, overall residual wage inequality will typically rise if there is a rising share of workers with above-average levels of within-group inequality.

Let $t_x = b$ denote the base year, for which the composition of the work force will be held fixed, and $t_w = o$ the year for which we intend to estimate a counterfactual wage distribution. We call this year the observation year. Here, we only use observations on full-timers in years t_w and t_x. The counterfactual wage distribution using the conditional wage structure of year $t_w = o$ but the distribution of characteristics x from the base year $t_x = b$ is given by

$$\begin{aligned} f(w|t_w = o, t_x = b) &= \int_x f(w|x, t_w = o) dF(x|t_x = b) \\ &= \int_x f(w|x, t_w = o)\rho(t_x = b) dF(x|t_x = o). \end{aligned} \tag{2.1}$$

where $f(w|t_w = o, t_x = o)$ is the actual density of wages for characteristics x in year $t_w = o$ and $\rho(t_x = b) = \frac{dF(x|t_x=b)}{dF(x|t_x=o)}$ is the reweighting factor which translates the density of observed wages into the counterfactual density. Note that as a special case $f(w|t_w = o, t_x = o) = \int_x f(w|x, t_w = o) dF(x|t_x = o)$, for which $\rho(t_x = b) \equiv 1$ in equation (2.1). The reweighting factor can be written as the ratio $\rho(t_x = b) = \frac{P(t=b|x)}{P(t=o|x)} \frac{P(t=o)}{P(t=b)}$, where $P(t = o)$ and $P(t = b)$ are the sample proportions of the observation year and the base year when pooling the data for both years.

The proportions $P(t = b|x)$ and $P(t = o|x)$ are estimated by logit regressions of the respective year indicator on flexible specifications of the characteristics shown in table 2.1. The logit regressions are based on the sample pooling the base year and the observation year. Using the fitted logit probabilities, we then calculate the individual reweighting factors $\rho_i(t_x = b)$ for observations i. All our estimates use the sample weights s_i which compensate for the varying length of employment spells. For robustness reasons, we trim the maximum value of individual observation weights to the value of thirty, in order to prevent extreme values of the reweighting factor, which may occur as a result of extremely rare combinations of characteristics. We tested a range of trimming thresholds, and found that values between 20 and 50 avoid extreme outliers, while at the same time excluding a very small number of observations (details are available upon request).

The reweighting factor can be incorporated into the estimation of counterfactual quantiles based on the sample wage distribution while fixing the composition of full-timers in the base year. Using the abbreviation $\rho = \rho(t_x = b)$, the reweighted (composition adjusted) p% quantile is given by

$$Q_p(w|t_w = o, t_x = b) = \begin{cases} \frac{w_{[j-1]}+w_{[j]}}{2} & \text{if } \sum_{i=1}^{j-1}(s\rho)_{[i]} = \frac{p}{100}\sum_{i=1}^{n}(s\rho)_{[i]} \\ w_{[j]} & \text{otherwise} \end{cases}, \tag{2.2}$$

where

$$j = min\left(k|\sum_{i=1}^{k}(s\rho)_{[i]} > \frac{p}{100}\sum_{i=1}^{n}(s\rho)_{[i]}\right),$$

$w_{[i]}$ is the ith order statistic of wages, and $(s\rho)_{[i]}$ is defined accordingly (i.e. the order statistic of the compound individual weights $s\rho$, combining the sample weight s with the reweighting factor ρ).

We consider the quantile gaps (differences in quantiles of log wages) between the 85th and 50th, the 85th and 15th as well as the 50th and 15th counterfactual percentile, i.e.

$$QG_{85/50}(w|t_w = o, t_x = b) = Q_{85}(w|t_w = o, t_x = b) - Q_{50}(w|t_w = o, t_x = b) \tag{2.3}$$
$$QG_{85/15}(w|t_w = o, t_x = b) = Q_{85}(w|t_w = o, t_x = b) - Q_{15}(w|t_w = o, t_x = b) \tag{2.4}$$
$$QG_{50/15}(w|t_w = o, t_x = b) = Q_{50}(w|t_w = o, t_x = b) - Q_{15}(w|t_w = o, t_x = b). \tag{2.5}$$

In addition to a graphical comparison of the actual and counterfactual development over time, we also contrast the increase in the counterfactual quantile gaps with the actual increase between 1985 and 2010. This allows us to quantify the share of the increase in inequality associated with composition changes (where $g \in \{85/50, 85/15, 50/15\}$)

$$shareQG_{g,x}(w|t_w = 2010, t_x = 1985) = \tag{2.6}$$
$$\frac{QG_g(w|t_w = 2010, t_x = 2010) - QG_g(w|t_w = 2010, t_x = 1985)}{QG_g(w|t_w = 2010, t_x = 2010) - QG_g(w|t_w = 1985, t_x = 1985)}.$$

For the logit regression, we use a sequence of specifications adding covariates in order to investigate the incremental composition effect on wage inequality. We divide the vector of characteristics into five groups of variables, namely educational outcomes (Ed), labor market experience (Ex), labor market history (Hist), occupation and industry characteristics (Occ, Ind) (see tables 2.1 and A2.14). Among those, we consider potential labor market experience

Table A2.14: Specification overview

Label	Covariates	Specific covariates
E	Education	ed
EE	Education, Experience	$ed, ex, ed * ex, ex^2, ed * ex^2$
EEH	Education, Experience, Labor market history	$ed, ex, ed * ex, ex^2, ed * ex^2, pt, ft, pt5, ft5, ed * pt5, ed * ft5, pt5^2, ft5^2, ed * pt5^2, ed * ft5^2$
EEHOI	Education, Experience, Labor market history, Occupation & Industry	$ed, ex, ed * ex, ex^2, ed * ex^2, pt, ft, pt5, ft5, ed * pt5, ed * ft5, pt5^2, ft5^2, ed * pt5^2, ed * ft5^2, occ, occ * ex, occ * ex^2, sec, sec * ex, sec * ex^2, sec * ed$

Note: Covariates used for reweighting procedure. E.g. $ed, ex, ed * ex, ex^2, ed * ex^2$ reads that education, experience, experience-squared and the two interactions education*experience and exducation*experience-squared are used.

as continuous and all other variables as categorial, leading to a highly flexible specification of the logit model. We calculate four versions of the counterfactual quantile gaps, starting with a specification only controlling for education (row E in table A2.14).

Sequentially adding sets of covariates (characteristics) to our reweighting procedure, we estimate the change in the counterfactual quantile gaps that is associated with the set of covariates considered so far. This way, we quantify the incremental contribution of covariatess to the rise in wage inequality (this contribution is given by the figures in the columns labeled 'Increment', see e.g. table A2.4). We decompose the difference between the observed and counterfactual rise in inequality into the effects of separate sets of covariates. For example, when adding occupation and industry characteristics (OI) to the reweighting function that already contains education, experience and labor market history (EEH), we measure the incremental effect of occupation and industry (OI) net of the effect contributed by the set of covariates already included (EEH). We add covariates in the order given in table A2.14. The incremental effect of each set of covariates depends upon the order in which they are added to the model. Our reasoning behind the choice of the sequence shown in table A2.14 is that we gradually move from exogenous and predetermined characteristics towards characteristics that are the likely consequence of endogenous decisions of the individual. We start with education because education typically remains fixed after labor market entry. Next, potential work experience is a linear function of time and education. Similary, labor market history involves characteristics which are affected by education and actual work experience. Finally, occupation and industry can in principle be changed any time conditional on education, experience and labor market history, and we are particularly interested as to whether occupation and industry play a role after accounting for all other individual level characteristics.

One may wonder how the reweighting method deals with endogeneity, i.e. unobservables that are not included in the analysis but that are potentially correlated with the included observables. Fortin et al. (2011) show that for a causal interpretation one only has to make the assumption that the distribution of unobservables for workers with identical observables (including observed labor market history) is the same in the base year and the target year (assumption 5, p. 21 in Fortin et al. 2011). Note that this does not rule out correlation of observables and unobservables. Put differently, the relationship between observables and unobservables is assumed to be time-invariant. This assumption would be violated, if e.g. having prior part-time/nonemployment experience is increasingly associated with good or bad unobservables. While we cannot rule out this possibility, there is no evidence for such an effect. However, the point to be stressed is that a mere correlation between observables and unobservables does not pose a problem to our method as long as the correlation does not vary systematically over time.

Composition reweighting for total employment

The reweighting can be expanded to take into account selection between full-time work and total employment based on observables, thus addressing the limitation that the SIAB data do not provide comparable wages for part-timers. We first calculate wage distributions for full-timers using the distribution of characteristics in the total employment sample, involving both part-timers and full-timers. Then, in a second step, we reweight these counterfactual wage distribution to the characteristics of a base year, analogous to section 2.6.4. The resulting distribution can be interpreted as the wages that would have prevailed had all individuals worked full-time and had their characteristics stayed at the level of the base year.

The first step consists in within-period composition reweighting. We calculate counterfactual wage distributions, which would have prevailed if all individuals had been paid full-time wages. This interpretation holds under the assumption that returns to characteristics for non-full-timers are equal to those for full-timers. The results of Manning and Petrongolo (2008) suggest that hourly wage differentials for (female) part-timers in industrialized countries are not driven by differences in returns to characteristics, which lends credibility to our approach. In order to calculate these distributions, we apply the reweighting technique described in section 2.4.1, but instead of the full-time sample in a specific base year, the reference group is total employment in the same year. Let $e \in \{FT, TE\}$ describe the employment group to which each observation belongs, where FT represents full-timers and TE total employment. Full-time workers appear in both FT and TE. The reweighting factor $\rho(FT \to TE, t_x = o)$ is the probability of characteristics x in the total employment sample in a given year, relative to the probability x in the full-time sample of the same year

$$\begin{aligned}\rho(FT \to TE, t_x = o) &= \frac{dF(x|e_x = TE, t_x = o)}{dF(x|e_x = FT, t_x = o)} \\ &= \frac{P(e = TE|x, t = o)}{P(e = FT|x, t = o)} \frac{P(e = FT|t = o)}{P(e = TE|t = o)}\end{aligned} \tag{2.7}$$

Then, the counterfactual distribution of wages, assuming the entire labor force was working full-time, can be written as

$$\begin{aligned}& f(w|e_w = FT, e_x = TE, t_w = o, t_x = o) \\ = & \int_x f(w|x, e_w = FT, t_w = o, t_x = o)\rho(FT \to TE, t_x = o)dF(x|e_x = FT, t_x = o)\end{aligned} \tag{2.8}$$

Here, $P(e = TE|x, t = o)$ is estimated by a weighted logit regression on the pooled sample of the reference group (total employment TE) and the group of interest (full-timers FT), with the employment status indicator e denoting group membership of each observation. In this step, we use the specification from table A2.15, in order to include the full set of observable individual characteristics.

Table A2.15: Specification for counterfactual total employment

Variables	Specific covariates
Education, Experience, Labor mar- ket history, Occupation, Industry	$ed, ex, ed*ex, ex^2, ed*ex^2, pt, ft, pt5, ft5, ed*pt5, ed*ft5, occ, occ*ex, occ*ex^2, sec, sec*ex, sec*ex^2, sec*ed$

Note: Covariates used for reweighting procedure.

In a second step, we analyze the distribution of wages which would have prevailed, had all employees worked full-time, and had their characteristics been fixed at the level of the base year. By holding the composition of total employment constant over time, we control for changes in the wage distribution due to changes in the selection into total employment over time. This counterfactual distribution can be written as

$$\begin{aligned}& f(w|e_w = FT, e_x = TE, t_w = o, t_x = b) \\ = & \int_x f(w|x, e_w = FT, t_w = o)\rho(e_x = TE, t_x = b)\rho(FT \to TE, t_x = o)dF(x|e_x = FT, t_x = o)\end{aligned} \tag{2.9}$$

where

$$\rho(e_x = TE, t_x = b) = \frac{dF(x|e_x = TE, t_x = b)}{dF(x|e_x = TE, t_x = o)} \tag{2.10}$$
$$= \frac{P(t = b|x, e_x = TE)}{P(t = o|x, e_x = TE)} \frac{P(t = o|e = TE)}{P(t = b|e = TE)}$$

Analogous to section 2.6.4, we sequentially add groups of covariates to our logit specifications as described by table A2.14. This allows us to investigate the incremental changes in inequality associated with the corresponding composition changes.

Choice of base year and interaction effects

As a robustness check and to account for interaction effects in the counterfactual analysis, we reverse the role of the base year and the target year in our reweighting procedure. So far, we have considered the wage distribution in 2010 and changed the distribution of characteristics back to that of the base year 1985. This is indicative of the part of the inequality increase that could be 'reversed' by undoing the change in characteristics. In this case, the inequality change explained by composition effects is $QG(t_w = 2010, t_x = 2010) - QG(t_w = 2010, t_x = 1985)$. Now, we focus on the opposite case in which we start with the wage distribution in 1985 but only change the distribution of characteristics to the level of 2010. This correspondends to the change $QG(t_w = 1985, t_x = 2010) - QG(t_w = 1985, t_x = 1985)$, i.e. the part of the inequality increase that can be accounted for by solely changing the distribution of characteristics while holding fixed the conditional wage structure of 1985.

Figures A2.19 to A2.22 and tables A2.12, A2.13 report the findings. For males, the contribution of the different sets of covariates to the overall inequality increase remain qualitatively similar, with a few notable exceptions. The general result is that compositional changes in educational qualifications and in labor market histories provide substantial contributions, while compositional changes related to potential work experience and the occupations/industry structure do so only to a much smaller extent (table A2.12 vs. table A2.4). However, the impact of education changes is much stronger in table A2.12 compared to table A2.4 (31.9%. 59.4%, 7.4%, 18.6% vs. 17.1%, 37.5%, -1.0%, 7.1%). This means that compositional changes over time are associated with a stronger rise in wage inequality based on the wage distribution in 1985 compared to 2010.[16] Put differently, the effects of

[16]This conclusion is based on the following formal argument (10≡2010,85≡1985):

$$QG(t_w = 85, t_x = 10) - QG(t_w = 85, t_x = 85) > QG(t_w = 10, t_x = 10) - QG(t_w = 10, t_x = 85)$$

a widening conditional wage structure $f(w|x)$ is stronger when applied to the distribution of characteristics in 1985 than when applied to that in 2010. This would naturally arise if the 1985 distribution of characteristics is more heterogeneous so that applying diverging wage returns to this more heterogeneous population leads to stronger inequality increases. Take education, the share of low-skilled declines from a high initial level, while the share of high-skilled increases (figure A2.5). Another difference between tables A2.4 and A2.12 is that the contribution of occupations/industries falls when the base year 2010 is used (table A2.12). In contrast to the results for education, the composition of occupation and industry has changed in a way that wage inequality increases more strongly for the 2010 composition of occupation and industry compared to the 1985 composition.

For females, the contribution of composition changes in work experience and recent labor market histories remains qualitatively unchanged when we change the base year (columns 6 to 10 in tables A2.5 and A2.13). As for males, the compositional effects of educational upgrading becomes much stronger in table A2.13. The only other effect for females, that is not fully robust to the choice of the base year, concerns the changes in occupations and industries. Here, table A2.13 shows pronounced effects on inequality in the upper and lower part of the distribution, which are not present in table A2.5. The overall contribution of compositional effects to rising female wage inequality in table A2.13 is even larger than for the base year 1985 (table A2.5). In particular, composition changes can account for 78.4% (103.2%) of the rise in female overall (lower tail) wage inequality between 1985 and 2010. We conclude that the composition changes would have been associated with a large increase in inequality based on 1985 wages compared to 2010 wages. This is in contrast to the widely held view in the past that Germany used to be a country where institutions strongly limited wage inequality (see Fitzenberger 1999 or Dustmann et al. 2014 for a critical assessment of this view).

is equivalent to

$$QG(t_w = 10, t_x = 10) - QG(t_w = 85, t_x = 10) < QG(t_w = 10, t_x = 85) - QG(t_w = 85, t_x = 85).$$

Chapter 3

Changing Selection into Full-time Work and its Effects on Wage Inequality – An Application to Germany

3.1 Introduction

Germany experienced a considerable increase in wage inequality until 2010 (Dustmann et al., 2009; Card et al., 2013; Möller, 2016; Biewen et al., 2018). For an assessment of what factors drive the observed changes in the wage distribution and the wage differences between labor market groups, it is necessary to take into account that the selection into paid work may change over time or may differ across groups. Importantly, selection into paid employment may depend on observed as well as on unobserved characteristics. Selection may work through the changing composition of the workforce with respect to easily observable characteristics, such as educational qualifications, work experience or age. It may also work through selection based on unobserved factors like motivation, social skills or the ability to adapt to changing circumstances. For example, increasing labor force participation is likely to draw individuals into the work force who differ with respect to their unobserved skills when compared to individuals who are already employed. The German labor market is subject to considerable employment fluctuations through unemployment and considerable long-term employment growth (see, for instance, Ljungqvist and Sargent (1998)). Given these developments, changes in selection and changing composition are important for understanding the evolution of the wage distribution over time and for a comparison of wage distributions between groups. For instance, the rise in wage inequality over time will be overestimated if falling unemployment draws people into employment who represent a negative selection of all workers. In this case, increasing wage inequality can be considered the sign of a positive development because it implies that individuals who did not have employment are entering work. Even if these individuals are earning wages at the lower end of the distribution, this is an improvement over being unemployed. This was the motivation underlying a range of reforms to labor market regulations between the late 90s and late 2000s in Germany. These reforms were aimed at increasing labor force participation and easing transitions from unemployment into low quality jobs. However, if individuals newly drawn into the labor force are not negatively selected with respect to unobserved qualities, rising wage inequality is a sign of increasing disparity in wages among individuals who were previously earning homogeneous wages.

This paper estimates selection-corrected quantile regressions to address two research questions regarding wage inequality among German men in 1995 and 2010. First, we consider the shape of the wage distribution and the magnitude of inequality in wages which would have prevailed if all unemployed had been working full-time. Because full-time employment is selective and likely based on earnings prospects, we would expect wage inequality to be higher if both the unemployed and the employed were working full-time. Our second question addresses the changes over time: How would wage inequality have developed if selection into full-time employment had not changed over time?

If the distribution of observed and unobserved skills was the same among unemployed and employed individuals, correcting for selection would not be an econometric challenge. However, we expect full-time workers to differ substantially from unemployed workers. Then for estimating the parameters of wage offer functions, standard regression techniques will usually be biased. Thus a common approach when analyzing the mean of the wage distribution is to apply sample selection corrections based on Heckman (1979). These approaches are not easily generalized to the entire distribution of wages, and therefore to measures of inequality such as quantiles. There exists a small but growing literature as to how to account for unobservables in the analysis of wage distributions. For instance, Card et al. (2013) estimate worker and firm fixed effects accounting for unobservable persistent differences between workers and between firms. However, the study does not account for the selection into employment due to unobservables.

A limited number of approaches have been suggested to correct entire distributions for selection due to unobservables. Most applications of selection corrected quantile regressions so far employ a control function approach, as in Buchinsky (2001; 1998), Albrecht et al. (2009), Bollinger et al. (2011) and Picchio and Mussida (2011), which we also apply for wage regressions based on German administrative data (see also Das et al. 2003 for semi-parametric selection models). Huber and Melly (2015) point out that this selection correction approach is only valid if the error terms in the selection equation and the wage equation are independent conditional on the selection probability. This conditional independence assumption implies equal slope coefficients for the determinants of wages in the selection corrected quantile regressions of wages.

As our methodological contribution, we propose to respecify the estimated selection corrected quantile regressions by transforming the dependent variable with the goal that equality of the slope coefficient then holds. The transformation is estimated based on the identification-at-infinity assumption which is plausible in our application and which ensures conditional independence in our application.[1] Our approach is a modification of the two-step approach by Buchinsky (1998) which includes an additional step to address the concern raised by Huber and Melly. A version of their test of equality of slope coefficients is used to guide the choice of the transformation. With the control-function approach augmented by a transformation of the dependent variable, we estimate quantile regressions which are corrected for selective movements between unemployment and full-time work. Undoing the transformation based on the selection corrected quantile coefficients and employing the decomposition technique of Melly (2005), we construct counterfactual wage distributions.

n a recent important paper, Arellano and Bonhomme (2017) suggest a copula based method to provide a consistent estimate of quantile regressions with selection correction. They estimate quantile regressions while assuming a fixed copula between the conditional rank in the wage distribution and the rank in the error term of the selection equation. The approach amounts to estimating rotated quantile regressions, which relate the τ^{th} quantile regression in the nonselected sample to a rotated value (the rank of τ^{th} unselected quantile) in the selected sample thus linking the two for estimation purposes. This is an alternative to Buchinsky's selection corrections approach which estimates the difference between the τ^{th} quantiles in the two samples. The approach of Arellano and Bonhomme (2017) has two disadvantages. First, the authors estimate the copula while assuming a specific functional form, and they allow only for the covariates to have a limited impact on the joint distribution of ranks. Second, the estimation of the copula is computationally very

[1] The idea to transform the dependent variable is similar to the approach applied in the companion paper Biewen et al. (2019), which estimates the selection bias in employment for the estimation of the gender wage gap. However, the actual implementation of the transformation approach differs between the two papers.

involved. D'Haultfoeuille et al. (2014) suggest an approach in which identification relies on the independence between covariates and selection for large values of the outcome, and on the homogeneity of the estimand across the distribution. This assumption does not seem plausible in our application.

Our findings show that the unemployed are a negative selection of the workforce, conditional on education, in times of low unemployment as in the years 1995 and 2010. The counterfactual wage distributions if everyone was working full-time would lie below the observed wage distribution and wage inequality would be much higher. This suggests that the unemployed would earn more heterogeneous wages than the already employed if working full time. Put differently, the counterfactual wage distributions if everyone was working full-time have thicker lower and upper tails than the observed wage distribution.

For our second research question, we find that those employed in 1995 would have had lower wages in 2010 then those employed in 2010 and wage dispersion would have been higher. Overall, this implies that full-time workers have become less heterogeneous with regard to the factors driving wages as well as the selection into full-time work.

The remainder of this paper is organized as follows: In section 3.2 we describe the data used and provide descriptive evidence of trends in wages, unemployment and the instrumental variables for the control function approach. Section 3.3 describes in detail our econometric approach for estimating selection corrected quantile regressions and calculating counterfactual wage distributions. We apply this approach to our data and discuss the results in section 3.4. Section 3.5 concludes.

3.2 Data and descriptive evidence

For our analysis we use the SIAB, an administrative dataset based on German social security records. This study uses the factually anonymous Sample of Integrated Labour Market Biographies (version 1975-2010, henceforth denoted by SIAB710).[2] It contains a 2 percent sample of all dependent employees who are subject to social security, all individuals receiving unemployment benefits, but no self-employed or civil servants. We restrict the analysis to those between the ages of 25 and 55 who are working in West-Germany. Wages are available as daily wages in Euros, which we deflate to the level of 1990 and take the log of. Since

[2]We used the Scientic Use File supplied by the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB), see Vom Berge et al. (2013) for the data documentation.

these wages are collected from administrative data sources, the measurements are very precise and do not suffer from the problems of selective nonresponse or measurement error commonly associated with wage information in survey data. While our dataset does not contain information on hours worked, we are confident that daily wages among full-time employees are sufficiently comparable. However, without working hours for part-time and marginally employed workers, wage data for those observations is not comparable across observations and jobs. Labor supply decision might vary greatly across time and between individuals, which would create strong confounding effects. We therefore restrict attention to full-time employees, thus following of literature (Dustmann et al., 2009; Card et al., 2013; Möller, 2016). As a consequence of the unavailability of comparable wages for part-time workers and because we do not observe individuals outside the labor force, we perform our analysis only on the data for males. For males, the majority of selective movements during working age occur between unemployment and full-time employment. However, this is not the case for females. For them, part-time employment and absence from the labor force affect large shares of the working age population, so any analysis which focuses on movements between unemployment and full-time employment is not well suited to studying the effects of selection on female wages.

We focus on the years 1995 and 2010. Those years represent the start and end of the greatest rise in wage inequality for German workers, as well as the turning point in the development of unemployment ((Biewen et al., 2018), (Möller, 2016)). Table A3.1 involves descriptive statistics on the samples used for our analysis.

Levels of education are aggregated into three categories based on highest degrees obtained: (i) High-educated: College (University/University of the Applied Sciences), (ii) Medium-educated: High school and/or Vocational Training, and (iii) Low-educated: No/Other degree. We capture each individual's labor market history as the number of days spent in full-time employment and part-time employment, aggregated over the last 5 years, respectively. Episodes of part-time and non-employment are important determinants of individual wage development, as shown by Paul (2016) and of changes in wage inequality in general (compare (Biewen et al., 2018). All wages above the contribution threshold for social security, which lies between the 85th and 90th percentile, are censored in the sample. For the analysis of wage quantiles above the threshold we impute wages, using a method based on Gartner (2005). The imputed wages are based on the fitted values of a Tobit model for censored data and take into account the heteroscedastic variance of the Tobit model. Because of the severe censoring for the high-educated, we restrict our analysis to the medium- and low-educated.

3.2.1 Wage inequality

From the early 1990s onward, wage inequality increased substantially, as measured for instance by the gap between the top quartile and the bottom quartile of the distribution of gross wages. The top left panel of Figure 3.1 shows that, relative to their levels in 1995, male workers near the bottom of the wage distribution suffered a decline in wages, while those near the top experienced a rise. The median wage has been basically stagnant over the entire period from 1995 to 2010, but workers in the lower half of the wage distribution are now earning substantially less in real terms than they did in the nineties. Some increase in inequality can be considered a natural result of diversified wages due to an aging population and increased shares of highly educated workers (Dustmann et al., 2009; Biewen et al., 2018). Generally though, the observed increase in inequality is seen as a negative development, because it reflects falling wages for low earners. This has caused great concern among policymakers and might have contributed to the introduction of a statutory minimum wage in Germany (Caliendo et al., 2019).

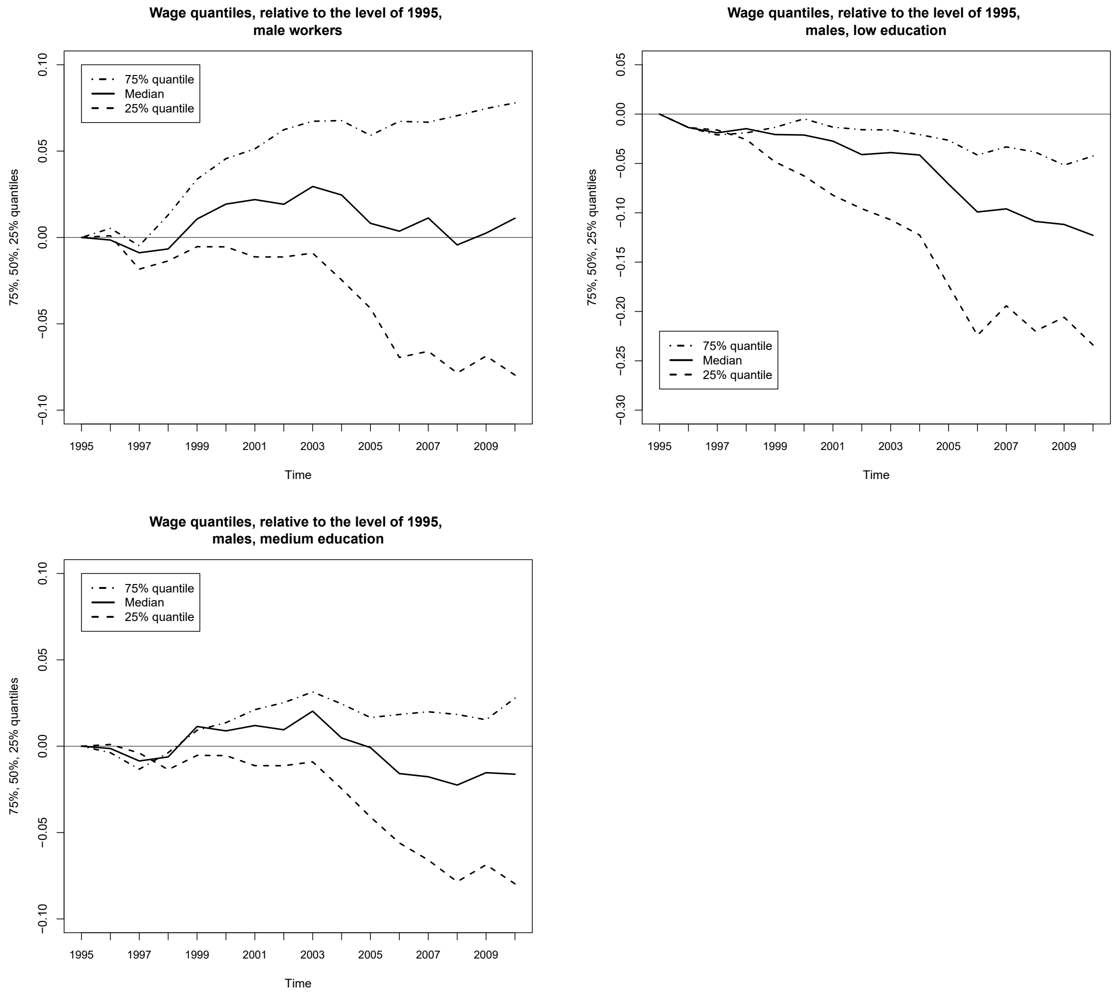
Even within education groups, the wage distributions have widened since the mid 1990s. As shown in panels 2 to 3 of Figure 3.1, this wage inequality increased strongly both for the low-educated and the medium-educated. For the low-educated, real wages fell even up to the top of the wage distribution, even above the upper quartile, and the decline of the median real wage between 1995 and 2010 amounts to about 10 log points.

3.2.2 Unemployment

We focus our analysis on the unemployed which are subject to unemployment insurance and are receiving unemployment benefits (Arbeitslosenhilfe, Arbeitslosengeld 1). Unemployment insurance benefits in Germany are paid for a maximum of 12 months to individuals who previously had a spell of dependent employment.[3] The registered unemployment rates for German men have changed substantially between 1995 and 2010. They start at 7.5 percent in 1995 and reach their peak of 9.8 percent in 2004. After 2004, we initially see a decline in unemployment which rises again slightly in the aftermath of the financial crisis. For the subgroup with medium education, the aggregate changes in unemployment rates are

[3]Unemployment benefits are paid for a longer time period above certain age limits, which applies mostly to workers above age 55. Long term unemployed are covered by other types of welfare which have undergone multiple reforms over the observation timeframe and are not consistently observed in the dataset. Additionally, not all of those receiving welfare benefits are available for employment (e.g. due to illness or early retirement with pensions below welfare levels). We therefore refrain from including the long-term unemployed in our analysis, as they are not well suited for analyzing counterfactual wages if employment was not selective with respect to worker characteristics.

Figure 3.1: Wage inequality

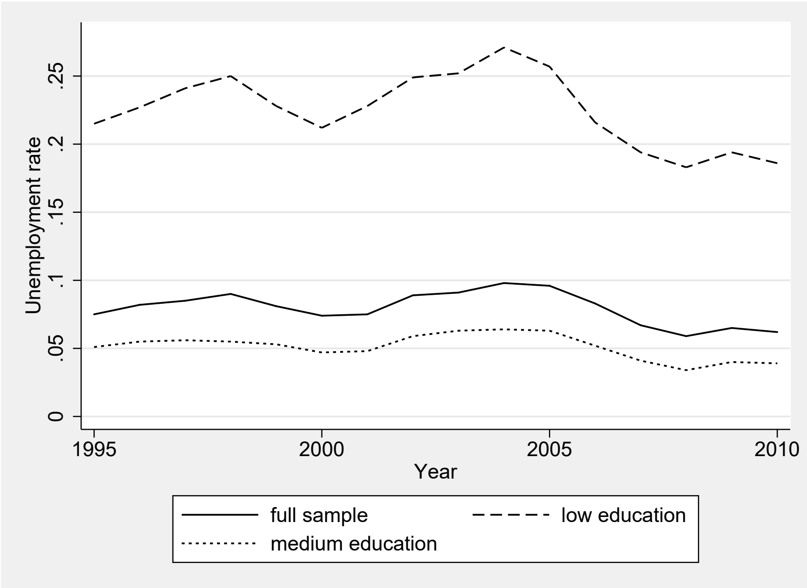


Note: Wage quantiles among full-time male workers in West-Germany, normalized to the levels of 1995, separately by education category. Source SIAB, own calculations.

closely mirrored. Among those with low education unemployment rates are generally higher, especially before 2005, but decline even more strongly than among the medium educated. The strong drop in unemployment between 2004 and 2010 coincides with the rapid increase in wage inequality documented in Section 3.2.1.

A common interpretation is that the fall in unemployment could be associated with a stronger inflow of previously unemployed into full-time work (see e.g. Dustmann et al., 2014). Those previously unemployed individuals might, on average, possess observable and unobservable characteristics which are less highly valued in the labor market than those of the previously employed workers. Therefore, the resulting labor force may be more hetero-

Figure 3.2: Unemployment rates



Note: Source IAB Labor market report 10/2017. Unemployment rate among male workers in West-Germany, separately by education category.

geneous with regard to the drivers of wages (Biewen et al., 2018). Because work incentive for low-wage workers have been strengthened by various labor market reforms in the early 2000s, this effect could be particularly strong in the lower tail of the wage distribution, contributing to the decline of the quantiles below the median.

However, it is an open question whether a decline of unemployment necessarily implies a widening of the lower tail of the wage distribution. We would like to mention three possible counter arguments without being able to provide a comprehensive discussion. First, labor market frictions might prevent wages of newly employed to differ substantially from those of the already employed. Second, the cuts in unemployment benefits may also have reduced the bargaining power of the incumbent workforce. Third, rising rates of retirement, which also reduce the number of older individuals in unemployment, a falling supply of younger workers, and higher wage flexibility among younger workers may reduce unemployment but not widen the wage distribution.

3.2.3 Instruments for selection

Semiparametric identification of selection effects in quantile regressions of wages requires at least one instrument satisfying an exclusion restriction (compare Buchinsky, 1998), analogous to a Heckman sample selection model for mean regression. The instruments need to provide exogenous variation of the selection probability into employment without affecting wages. Since the SIAB7510 data do not contain individual level variables, which we think are suitable as instruments, we use as instruments four additional variables merged to the SIAB7510 at the regional district level (Kreisebene). These variables are cohort sizes of young adults aged 18 to 25 and 25 to 30 as well as graduation rates in lower secondary and higher secondary education. These instruments reflect exogenous shocks to the labor supply in the respective region and year, affecting individual employment chances. We believe the exclusion restrictions to be credible, because it is unlikely that wages respond in the short run to labor supply changes between regional districts. Wage rigidities prevent short term adjustment in response to labor supply variations due to new entrants into the labor market (compare Bauer et al., 2007). This is partly because wage contracts generally span multiple years and wages of new employees are not independent of wages for current employees, after accounting for individual differences in employment history. Additionally, collective bargaining in Germany work at the level of the industry or large firms and therefore do not allow for a wage response to shocks at the district level. District level data on the instruments are obtained from the Federal Statistics Office's regional database.[4] Our analysis will rely on an identification-at-infinity assumption, meaning that the support of the instrument includes with positive probability cases, for which the selection probability is close to one (Heckman, 1990a).

3.3 Methodological approach

3.3.1 Model setup

The model setup follows the notation of Huber and Melly (2015). The wage equation for all individuals (employed or unemployed) is

$$Y^* = X\beta + v \,, \tag{3.1}$$

where Y^* denotes the latent log wage in the absence of selection, X the vector of observable covariates, being determinants of wages, v the error term, and β the vector of coefficients. We assume that $\beta_{0.5} = \beta$, i.e. β represents the median coefficients and v represents the

[4]Data source: Regionaldatenbank des Statistischen Bundesamtes. If we include individuals aged 20 to 60, the strength of the instruments increases. This means that individuals in their early 20s and late 50s are more strongly affected by labor supply shocks of young workers entering the labor force. However, we restrict our empirical analysis to those 25 to 55 years old because different nonemployment states can not be distinguished well in our data. Many individuals aged between 20 and 24 are still in education and individuals in their late 50s start leaving the labor force through early retirement.

residual of a median regression. Assuming a linear quantile regression, the conditional τ-quantile of the latent wage $Q_\tau(Y^*|X)$ is specified by

$$Q_\tau(Y^*|X) = X\beta + Q_\tau(v|X) = X\beta_\tau \ , \tag{3.2}$$

which also means that $Q_\tau(v|X) = X(\beta_\tau - \beta)$ is a linear function of X. Correspondingly, the τth quantile regression of Y^* is $X\beta_\tau + v_\tau$, with $v_\tau = v - Q_\tau(v|X) = v - X(\beta_\tau - \beta)$.

The selection problem arises because we only observe wages for employed individuals. Let Y denote the observed wage and D the selection indicator. We specify

$$D = 1(Z\gamma + \varepsilon \geq 0) \ ,$$

where Z is a strict superset of X, thus also including instruments for selection, which are excluded in eq. (3.1), and ε is assumed to be independent of Z. The probability of selection

$$Pr(D = 1|Z) = Pr(Z\gamma + \varepsilon > 0|Z) \tag{3.3}$$

is a function of $Z\gamma$. For the selective sample, the observation rule is $Y = Y^*$ (Y^* observed) only if $D = 1$. A conditional quantile in the selected sample is

$$Q_\tau(Y|Z) \quad = \quad X\beta_\tau + Q_\tau(v_\tau|Z, D = 1) \ . \tag{3.4}$$

The term $Q_\tau(v_\tau|Z, D = 1)$ denotes the quantile-τ-specific selection bias, with $Q_\tau(v_\tau|Z, D = 1) > (<) 0$ representing positive (negative) selection. The selection bias can be rewritten as

$$Q_\tau(Y|Z) = Q_\tau(Y^*|Z, D = 1) = X\beta_\tau + \tilde{g}(X, Z\gamma) \tag{3.5}$$

where $Q_\tau(v_\tau|Z, D = 1) = \tilde{g}(X, Z\gamma)$ because v_τ depends on X and $D = 1$ on $Z\gamma$.

The control function $\tilde{g}(X, Z\gamma)$, which properly accounts for selection bias, should be a flexible function of X and $Z\gamma$, which is challenging because of the curse-of-dimensionality regarding X being multivariate. Nonparametric identification requires both independent variation of $Z\gamma$ given X and identification at infinity. Identification at infinity means that with positive probability, based on the distribution of $Z\gamma$, the selection probabilities $Pr(D = 1|Z)$ is close to one (Das et al., 2003). The selection model above implies that $Q_\tau(v_\tau|Z, D = 1)$ converges to zero (no selection), if the employment probability $P(D = 1|Z)$ converges to one, which is equivalent to $Z\gamma$ going to infinity.

Extending upon Heckman (1990a) and Andrews and Schafgans (1998), who consider the case where u is independent of X, both the intercept and the slope coefficients β can be identified, if we have observations with a selection probability close to one for each value of X. Given the linear specification of $X\beta_\tau$, a smaller subspace of the support A of X suffices, where $[E(X'X) \cdot I(X \in A)]$ can be inverted [$I(.)$ denotes the indicator function] and where the selection probability is close to one with positive probability. In our application, the selection probability is quite large for most observations and the subset of observations with a selection probability close to one (to anticipate: the median (upper quartile) of the selection probabilities lies above 93% (96%) in all four subsamples considered in our application, see table 3.1), is sufficiently large to estimate β_τ consistently. In our application, we will use the coefficient estimates based on the identification-at-infinity sample to characterize the selection bias in the full sample.

3.3.2 Buchinsky approach

The selection correction approach proposed by Buchinsky (1998; 2001) applies a standard Heckman selection approach with instruments (Heckman, 1979; Vella, 1998) to quantile regression. Buchinsky specifies the selection correction term in the second stage [eq. (3.3)] as a function of the inverse Mills ratio $\lambda = \frac{\phi(Z\hat{\gamma})}{\Phi(Z\hat{\gamma})}$, where ϕ and Φ are the density and the distribution function of a standard normal. However, even under joint normality of ε and v, the selection correction term $Q_\tau(v_\tau|Z, D = 1)$ is generally not a linear function in λ. Thus, Buchinsky suggests to approximate the selection correction term $Q_\tau(v_\tau|Z, D = 1)$ by a power series (polynomial) of λ (see Vella 1998 on semiparametric approaches for selection correction in mean regressions). Further, Buchinsky assumes that the joint distribution of v and ε is independent of Z, conditional on the probability of selection $Pr(Z\gamma + \varepsilon > 0|Z)$ (Huber and Melly, 2015).

In the second step, the selection corrected quantile regression

$$Q_\tau(Y|X) \quad = \quad X\beta_\tau + \theta_\tau g(\lambda) \tag{3.6}$$

is estimated for the selective sample with $D = 1$. Eq. (3.6) presumes that $\theta_\tau g(\lambda)$ represents $Q_\tau(v_\tau|Z, D = 1)$. $g(.)$ is a power series of λ, and thus $\theta_\tau g(\lambda)$ approximates the selection correction term $Q_\tau(v_\tau|Z, D = 1)$.

Without the assumption that the joint distribution of v and ε is independent of X conditional on $Z\gamma$, the selection model specified by eq.'s (3.2) and (3.3) implies that the selection correction term $Q_\tau(v_\tau|Z, D = 1)$ is some unknown function of both X and $Z\gamma$, see discussion of eq. (3.5) in section 3.3.1.

3.3.3 Huber-Melly test for conditional independence

Huber and Melly (2015) propose a quantile regression based test for the conditional independence assumption, stating that the joint density of v and ε is independent of Z conditional on $Z\gamma$. As noted by Huber and Melly (2015), Buchinsky's approach builds upon this conditional independence assumption, which implies homogeneous slope coefficients across all quantiles, see discussion of eq. (3.2) in section 3.3.1.[5]

We illustrate this point in the following. Conditional independence implies for the joint density of v and ε

$$f_{v,\varepsilon}(\cdot|Z) = f_{v,\varepsilon}(\cdot|Pr(D=1|Z)) = f_{v,\varepsilon}(\cdot|Z\gamma)\,. \tag{3.7}$$

When there is no sample selection, i.e. $Pr(D=1|Z) = 1\,\forall Z$, eq. (3.7) implies that v and ε are independent of Z. Under conditional independence, the quantile regression coefficients β_τ are identified when controlling for the selection bias term $Q_\tau(v_\tau|Z, D=1)$ as a flexible function of $Z\gamma$ only as in Buchinsky (1998, 2001), see also Huber and Melly (2015, section 2.2).

Conditional independence in eq. (3.7) also holds for v_τ and ε, implying that $Q_\tau(v_\tau|Pr(D=1|Z), D=1) - Q_\tau(v|Pr(D=1|Z), D=1)$ does not depend upon Z conditional upon the selection probability. Thus, the term $X(\beta_\tau - \beta)$ only involves a constant difference in the intercept, meaning that the slope coefficients in β_τ do not depend upon τ.

When the conditional independence assumption does not hold, slope coefficients β_τ may vary across quantiles, which is typically a motivation as to why researchers apply quantile regression in the first place. Huber and Melly (2015) point out that this limits the applicability of the Buchinsky approach to correct for sample selection in estimating quantile regressions.

Huber and Melly (2015) suggest a test based on the entire process of quantile regression coefficients to investigate whether the conditional independence assumption holds. They estimate quantile coefficients for a fine grid of quantiles across the distribution and then test the null hypothesis that the slope coefficients are identical. Violations of the null hypothesis are detected by using Kolmogorov-Smirnov (KS) and Cramér-von Mises (CM) test statistics to the coefficient process across quantiles. In practice, Huber and Melly use a grid of quan-

[5] The conditional independence assumption is implied by Assumptions C and E in Buchinsky (1998).

tiles grid and suggest to implement the test for a range from the 10th to the 90th percentile as a starting point. The first stage is estimated using the semiparametric Klein and Spady (1993) estimator. The sample selection correction is based on a polynomial in the inverse mills ratio of the estimated index function estimated. Inference is based on resampling the influence function of the quantile regression estimator, building on the differentiability of the selection correction function to take account of the first stage estimation error.

3.3.4 Our approach

In short, we first implement Buchinsky's approach based on the original data and then apply the Huber-Melly test which strongly rejects conditional independence. This is why we suggest to transform the dependent variable to account of heteroscedasticity in the original data and then apply Buchinsky's approach on the transformed dependent variable. Relying on identification-at-infinity, the transformation is based on quantile regressions for the subsample with a very high probability of participating. In our application, we are successful in finding a transformation for which Huber-Melly test passes afterwards. Note that this is not guaranteed and we undertake a specification search to find a proper transformation. If the conditional independence assumption is not rejected for the transformed model, we can use the transformed model to account for selection bias. Transforming back the dependent variable allows us to estimate counterfactual distributions in absence of selection or in the presence of a different selection mechanism.[6]

Now we describe in detail the different steps of our approach:

1. To estimate the probability to be in the selective sample, we estimate a Probit regression $Pr(D = 1|Z) = \Phi(Z\gamma)$, assuming that the distribution of ε in eq. (3.3) is independent of Z.[7]

[6]The basic idea to transform the dependent variable is similar to the companion paper Biewen et al. (2019), which estimates the selection bias in employment for the estimation of the gender wage gap. However, there are two key methodological differences. First, our paper only transforms the dependent variable while leaving the covariates unchanged, while Biewen et al. (2019) transform the both the dependent variable and the covariates. Second, our approach to determine the transformation factor relies on the identification-at-infinity approach, which is plausible in our setting. Biewen et al. (2019) assume a location-scale model $Y^* = X\beta + g(x)u$, where u is the rank in the conditional distribution of Y^* given x and derive a transformation factor based on the estimated conditional dispersion in the selective sample under the assumption that the dispersion of ranks in the selective sample is a function of the first stage selection probability.

[7]Huber and Melly (2015) use the alternative semiparametric estimator suggested by Klein and Spady (1993), which is also part of the code for the test provided by Huber and Melly (2015). We have experimented with both approaches (Probit and Klein and Spady) for some cases in our application and find little difference between the two with regard to the fitted probabilities. For simplicity and for computational reasons, the empirical analysis in this paper is based on the probit regressions for the first stage.

2. Based on the Probit estimates in step 1), a subsample of the data is determined for which identification-at-infinity is plausible. We estimate standard quantile regressions for these subsamples for which we assume that selection is negligible. Using coefficient estimates δ_u, δ_l at the upper quantile u and the lower quantile l, respectively, we then estimate the predicted conditional quantile differences (l and u are tuning parameters)

$$\sigma(X, \delta) = X\delta_u - X\delta_l \tag{3.8}$$

for a worker with characteristics X. The transformation then involves dividing Y by $\sigma(X, \delta)$.[8]

3. Next, we run selection corrected quantile regressions for the transformed outcome:

$$Q_\tau \left(\frac{Y}{\sigma(X, \delta)} \middle| X \right) = X\check{\beta}_\tau + g(\theta_\tau, Z\gamma) \,. \tag{3.9}$$

We specify the selection correction as a piecewise constant function, with $g(\theta_\tau, Z\gamma) = \sum_{j=1}^{4} \theta_{\tau,j} I(Z\gamma \in Q_j)$ involving dummies for four quintiles of the propensity score $I(Z\gamma \in Q_j)$ and $\theta_\tau = (\theta_{\tau,j})_{j=1,\ldots,4}$ (the highest quintile Q_5 represents the omitted category).[9] Then, as our implementation of the Huber-Melly test for conditional independence, we implement a Wald test of the equality of the slope coefficients $\check{\beta}_\tau$ along a grid of τ.

4. This step assumes that the conditional independence test in the previous step passes. We run OLS for the transformed model for the identication-at-infinity sample and then estimate the selection effect based on quantile regressions of the OLS residuals based on the entire sample. Compared to estimating quantile regressions for the transformed model with selection correction based on the full sample, the OLS estimator allows us to realize efficiency gains based on the high probability sample and then use the implied residuals based on entire sample to estimate the selection effects along the distribution.

5. Finally, we undo the transformation by multiplying the coefficients with $\sigma(X, \delta)$.

For simplicity, we implement the conditional independence test as a Wald test of the equal-

[8]This is analogous to the heteroscedasticity correction approach of Chen and Khan (2003), using a heteroscedasticity correction based on the inter-quartile range of the conditional distribution.

[9]This specification yields better fits and more reliable findings than using a polynomial in the inverse Mills' Ratio λ.

ity of slope coefficients over an equispaced grid of quantiles. Our application differs from Huber and Melly (2015) regarding the following three issues, which prevent us from using their code. First, bootstrapping the entire estimation process, inference takes account of the estimation error in all stages including the transformation. Second, applying a weighted cluster-bootstrap inference avoids nonconvergence of the probit in the first stage and is cluster-robust at the regional level, which is the level of the variation of the instruments.[10] Third, we approximate the selection correction term by a piece-wise constant selection correction function which is non-differentiable. Furthermore, implementing the Huber-Melly test for Buchinsky's estimator using a polynomial in the inverse-Mills-ratio based on the untransformed model requires a lot of computation time due to our large sample size.

If the conditional independence test for the tranformed model rejects, we use this for re-specifying our estimation approach. Note as a caveat that inference for our Wald tests for homegeneous slopes does not take account of the fact that we search for a transformation such that the conditional independence test passes. As part of our specifiction search, we investigate which quantile regression coefficients change strongly across quantiles. To illustrate this point, note that, based on prelimary estimates, the Huber-Melly tests never passed for a model pooling both education groups. Therefore, we conclude that the nature of the selection bias differs between the two education groups, which motivates us to estimate separate models by education group.[11]

3.3.5 Counterfactual wage distribution under alternative selection rules

We use the estimated selection corrected quantile regression to estimate the counterfactual wage distribution under different selection rules. We estimate the counterfactual distribution using a selection corrected Melly (2006) approach as in Albrecht et al. (2009) (see also Machado and Mata 2005; Chernozhukov et al. 2013), while taking into account the transformation of the outcome. Let Z, X, $g(Z\gamma)$ apply to the observed sample and $\tilde{Z}$, $\tilde{X}$, and $g(\tilde{Z}\tilde{\gamma})$ to the counterfactual sample, where $\tilde{\gamma}$ represents the counterfactual selection rule. Specifically, we estimate two counterfactuals: First, the wage distribution if all individuals in the sample were employed, and, second, the wage distribution if the selection rule of a different calendar year applies. The first counterfactual involves the covariates $\tilde{X}$ of the entire sample and sets $g(\theta_\tau, \tilde{Z}\tilde{\gamma})$ [i.e. $\theta_\tau = 0$] equal to zero, corresponding to a selection probability of one. For the second counterfactual, $\tilde{Z}$ and $\tilde{X}$ represent the employees and $g(\tilde{Z}\tilde{\gamma})$ their

[10]The code provided by Huber and Melly (2015) could be adjusted to provide cluster robust inference.

[11]Also, Machado (2017) finds differences in direction of selection across different sociodemographic groups.

selection rule (implied by the first stage Probit estimates) in the different calendar year.

Our implementation of the Melly (2006) approach uses predictions of conditional quantiles for a fine grid of equispaced $\tau \in [0.01, 0.02, ..., 0.99]$ for each observation in the counterfactual sample to estimate the conditional distribution of log wages. The counterfactual conditional quantile is

$$Q_\tau(Y|\tilde{Z}) = \sigma(\tilde{X}, \delta)\left[\tilde{X}\check{\beta}_\tau + g(\theta_\tau, \tilde{Z}\tilde{\gamma})\right] ,$$

where $\check{\beta}_\tau$, δ, and $g(\theta_\tau, .)$ (including the definition of the quintile dummies) are estimates based on the observed sample.

We then stack the 99 predictions for all individual observations in the counterfactual sample represented by $(\tilde{Z}, \tilde{X})$ and calculate the unconditional empirical quantiles of the entire expanded sample, where the number of observations is 99 times the number of observations in the counterfactual sample. This counterfactual distribution, denoted by $T_Y(\tilde{X}, \check{\beta}, \delta, \theta, \tilde{\gamma})$ represents the counterfactual distribution of Y for the sample with characteristics $\tilde{Z}$, the alternative selection rule $\tilde{\gamma}$, and the selection corrected coefficients for the transformed model $\check{\beta}$, the coefficients of the selection correction terms θ, and the transformation coefficients δ.

The difference between the counterfactual distribution $T_Y(\tilde{Z}, \check{\beta}, \delta, \theta, \tilde{\gamma})$ and the observed wage distribution (i.e. TO_Y representing the quantiles of Y in the selective observed sample with $D = 1$) describes the total effect of selection relative to the counterfactual, defined as

$$TO_Y - T_Y(\tilde{Z}, \check{\beta}, \delta, \theta, \tilde{\gamma}) . \tag{3.10}$$

We can now decompose the total selection effect into a component due to differences in observed characteristics driving wages, i.e. the difference between X and $\tilde{X}$, and a component due to differences in selection based on unobservables. To this end, we calculate a second counterfactual distribution based on linear quantile regression based on X in the observed sample of employees (without transformation) and then predicting the counterfactual distribution for the sample with $\tilde{X}$ using the Melly (2006) approach as described above. This counterfactual distribution is denoted by $T_Y(\tilde{X}, \alpha)$ where α involves the quantile regression coefficients for the observed sample.

The total selection effect in eq. (3.10) can be decomposed into the effect of changes in observable characteristics

$$TO_Y - T_Y(\tilde{X}, \alpha) , \tag{3.11}$$

and the residual effect of selection on unobservables

$$T_Y(\tilde{X},\alpha) - T_Y(\tilde{Z},\check{\beta},\delta,\theta,\tilde{\gamma})\,. \tag{3.12}$$

We now discuss the two cases separately, thereby defining the decompositions estimated in our empirical analysis

The first counterfactual wage distribution which would prevail if all observed individuals in a given year, both full-time workers and unemployed, were employed and earning market wages is obtained by setting θ_τ equal to zero. Then, eq. (3.10) defines the total effect of selection into work, which is decomposed into the selection effect due to observables [eq. (3.11)] and the effect of selection on unobservables [eq. (3.12)] when contrasting full-time workers with the total sample of full-time workers and unemployed.

The second counterfactual wage distribution allows us to study the effect of changes in selection over time. To estimate this counterfactual, we keep the conditional probability of selection into full-time work, i.e. the index $Z\gamma$, and the distribution of observed characteristics fixed at the level of the base year. Using the coefficient estimates obtained in the observation year (in our application the year 2010), we estimate the counterfactual wage distribution under the selection rule of a base year (in our application the year 1995). Let the index b denote the base year and o the observation year.

Then,

$$TO_Y^o - T_Y(Z^b,\check{\beta}^o,\delta^o,\theta^o,\tilde{\gamma}^b) \tag{3.13}$$

is the total selection effect. It can be decomposed as above into the effect of the change between base year and observation year in the selection of observables [eq. (3.11)] and in the selection on unobservables [eq. (3.12)], both among full-time workers.

Note the following caveat: These counterfactual distributions do not account for general equilibrium effects which might potentially lead to changing returns to skills in response to an influx of previously unemployed into employment (see the detailed discussion in (Fortin et al., 2011)). One likely response to such an influx would be falling returns to those skill levels over-represented among the unemployed, e.g. low levels of education. Therefore, returns to education might increase due to higher relative scarcity. Then, the estimated counterfactual wage distribution would be less dispersed than the one arising when all unemployed are employed and general equilibrium effects operate.

3.4 Application

3.4.1 First stage

Our decomposition method with sample selection correction requires instruments which affect the employment status but which do not affect wages. We run separate Probit regressions of the full-time indicator by education group, i.e. separately for the low-educated and the medium-educated.[12]

For the medium-educated, the probit regression accounts for the following covariates, which are also allowed to affect wages: Age, age squared, number of days spent in full-time work over the last 5 years, and number of days spent in part-time work over the last 5 years. As additional instruments, which are measured at the district level and which are excluded in the wage equation, we account for share of lower secondary graduates, share of upper secondary graduates, share of individuals aged 18-25, and share of individuals aged 25-30 in the district. The employment history variable account for the recent employment experience being associated with current full-time employment, thus accounting either for state dependence or for unobserved heterogeneity causing persistence in employment outcomes. Later these covariates are also used as control variables in the wage regression accounting for experience effects. We use labor supply instruments at the district level, assuming that wages are not affected by these supply instruments in the short run. Because we account for recent employment experience both in the selection equation and in the wage equation, this is compatible with labor supply changes affecting wages in the medium run. All covariates in the selection model are highly predictive for full-time employment among the medium-educated and have the expected signs (see table A3.2, columns 2 and 3). In particular, the excluded instruments are highly significant with an F-statistic of 20.4 in 1995 and 29.7 in 2010.

We also estimated the same specification for the low-skilled, however, the instruments were nowhere close to be significant.[13] Because the medium-educated are the larger group and the medium-educated may complement low-educated workers, we use the average fitted em-

[12]This is done for two reasons. First, the propensity scores based on a Probit regression pooling the two education groups and using the same regional instruments differ notably from those based on Probit regressions by education group. Second, we could not find a transformed model passing the Huber-Melly test when we account for selection based on a pooled Probit. Detailed results are available upon request. We conclude that the selection into full-time work and the effect of selection on observed wages differ between the two education groups.

[13]Note that in a specification pooling both education groups the instruments are significant. Recall, however, that pooling was rejected by the data.

ployment rate of the medium-educated at the district level based on the estimated selection equation in table A3.2, columns 2 and 3, respectively, as alternative instrument for the selection equation of the low-educated. The results for the low-educated are reported in table A3.2, columns 4 and 5. This instrument proves highly significant with an F-statistic of 11.7 in 1995 and 32.9 in 2010, implying that a higher employment rate of the medium-educated induced by labor supply changes also increases the employment rate of the low-educated.

As discussed in section 3.3.4, identification at infinity of the constant in the outcome model requires that the selective sample of the employed contains observations with a propensity score close to one. Figures 3.3 to 3.4 show that the distribution of the propensity score for the sample of employed and unemployed is concentrated close to one in all cases. Table 3.1 shows selected quantiles of the distribution of propensity scores for the selective sample of the employed. For the medium-educated, the median is 97% (97%) and the lower quartile is 95% (96%) in 1995 (2010). For the low-educated, the median is 94% (96%) and the lower quartile is 87% (78%) in 1995 (2010). Hence, we conclude that the identification-at-infinity approach described above is quite plausible for our application.

Table 3.1: Probability of selection among employed

	Medium-educated			Low-educated		
Year	25% quantile	Median	75% quantile	25% quantile	Median	75% quantile
1995	.945	.965	.977	.868	.936	.964
2010	.962	.970	.977	.778	.957	.967

Notes: Median and quartiles of the propensity score distribution restricted to the subset of full-time workers.

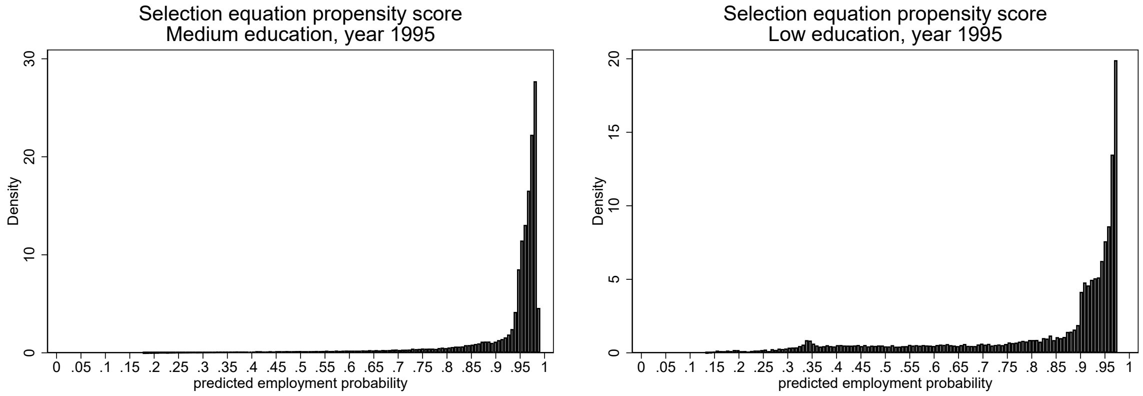
3.4.2 Conditional independence test for Buchinsky approach

We estimate Buchinsky's approach without transformation for selection corrected quantile regressions for our four cases using dummies for the quintiles of the propensity score to account for selection. Predicting the observed wage distribution in the employment sample using the Melly (2006) approach yields a close correspondence between the model prediction and the actual distribution.[14]

We then implement our version of the Huber-Melly test of equal slope coefficients β_τ for the selection corrected quantile regressions in eq. (3.6). We perform a grid search over a wide range of δ_u and δ_l for the second step transformation, to find the specification which

[14] In contrast, using a low order polynomial in the inverse Mills ratio did not result in satisfactory within sample fit.

Figure 3.3: Distribution of propensity scores in full sample by education group, year 1995



Notes: Propensity scores for being selected into full-time employment for the full sample of both employed and unemployed individuals based on estimates in table A3.2. Source SIAB7510, own calculations.

most effectively reduces differences in slope coefficients. Selected Wald-tests are reported in table 3.2. For the test range 80-20 ($\tau = .2, ..., .8$) the test statistics decisively reject in all cases. This also happens for the narrower test range 60-40 when implementing the test for all covariates. Only for the covariate part-time during the last 5 years, the test does not reject for the narrower test range. The rejection for all covariates is robust to other test ranges in between.[15] We conclude that Buchinsky's approach based on quantile regressions for log wages is not applicable for our application.

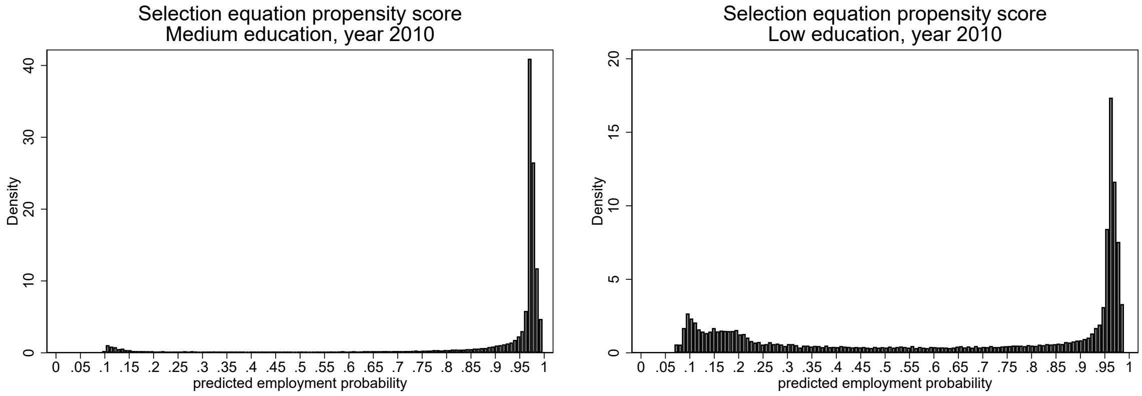
3.4.3 Transformation and conditional independence test - Steps 2 & 3

We use an identification-at-infinity sample to estimate the transformation factor $\sigma(X, \delta)$ in step 2 of our approach. For this, we use observations with a predicted probability above 90%/85% in 1995/2010 for the low-educated and above 97.5%/98% in 1995/2010 for the medium-educated, respectively. Based on different choices for the quantile range used for the transformation, we estimate quantile regressions with selection corrections as in step 3. Then we undertake the conditional independence tests and chose the transformation factor, i.e. the choice of δ_l, δ_u for the quantile differences used, according to the test results. The findings are reported in table 3.3 for our preferred models passing the conditional independence test.

In all cases, the conditional independence test passes for the narrow range 60-40 $[u - l =$

[15] Detailed results are available upon request.

Figure 3.4: Distribution of propensity scores in full sample by education group, year 2010



Notes: Propensity scores for being selected into full-time employment for the full sample of both employed and unemployed individuals based on estimates in table A3.2. Source SIAB7510, own calculations.

60% − 40%] and for all individual covariates for both reported ranges. For the medium-educated, the test passes for all covariates for 70-30 and also in 1995 for 80-20. For the low-educated, the test passes for 70-30 in 2010 and barely so at a 3%-level in 1995. One has to note that there are a three clear rejection for 80-20 considering all covariates, even though for the invidual covariates the test passes in all cases. Note that Huber and Melly (2015) caution themselves regarding the behavior of their conditional independence tests when moving into the tail of the distribution. The comparison between tables 3.2 and 3.3 shows that the transformation does a good job in reducing the differences in slope coefficients. We conclude that the conditional independence assumption is sufficiently plausible for the transformed model and the evidence for this is stronger for the medium-educated than for the low-educated.[16] With this in mind, we will be very cautious in not over-interpreting our selection findings.

3.4.4 Goodness-of-Fit and impact of selection - Steps 4 & 5

Assuming that conditional independence holds, we run OLS regressions without selection correction on the identification-at-infinity sample after the transformation. Then, we calculate residuals for the entire employment sample based on the OLS coefficient estimates. For these residuals, we then run quantile regressions on an intercept and the selection correction terms. Assuming that conditional independence holds, this focuses on the evolution of the

[16] Note that for quantile regressions after transformation pooling low-educated and medium-educated the conditional independence test is nowhere close to pass, i.e. the conditional independence test is not guaranteed to pass after a mechanical transformation. Detailed results are available upon request.

Table 3.2: Conditional independence tests for equality of slope coefficients in selection corrected quantile regressions without transformation (P-Values)

		Medium-Educated		Low-Educated	
Covariates	Test range for τ	1995	2010	1995	2010
All	80-20	.000	.000	.000	.000
All	70-30	.000	.000	.000	.000
All	60-40	.000	.000	.000	.000
Age + Age squared	80-20	.000	.000	.000	.000
Age + Age squared	60-40	.000	.000	.000	.000
Part-time 5years	80-20	.000	.000	.000	.000
Part-time 5years	60-40	.863	.278	.001	.000
Full-time 5years	80-20	.000	.000	.000	.000
Full-time 5years	60-40	.000	.000	.000	.000

Notes: P-values of specification tests under the null-hypothesis of conditional independence as in Huber and Melly (2015), testing for equality of the slope coefficients β_τ in eq. (3.6) over τ (e.g. range 80-20 denotes $\tau = .2, .25, ..., .75, .8$). Selection corrected quantile regressions as suggested by Buchinsky (1998). Wald tests on an equi-spaced five-percent-grid over the stated range for τ of the conditional distribution.

selection effects along the conditional distribution. Adding the OLS-fitted values to the fitted values of the quantile regressions for the residuals provide the quantile regression fits for the transformed model, which then can be used to simulate the wage distribution for the employed as well as the counterfactual wage distribution if all unemployed were also employed. These simulations are based on the Melly (2006) approach.

Contrasting the actual and simulated wage distribution for the employed allows us to assess the goodness-of-fit for the observed unconditional wage distribution. Figure 3.5 show that the fitted distributions closely track the actual distribution. Note that this is by no means obvious in light of our complex multi-step estimation approach. If the identification-at-infinity assumption were inappropriate or the transformation model/the model estimated for the transformed data were misspecified, the fitted distributions could differ from the actual distribution. The close fit between the actual and the fitted wage distribution also adds credibility to the estimated counterfactual distributions discussed below. Note that figure 3.5 shows the rise in wage inequality from 1995 to 2010. The 90-10 differential increases by about 15 log points for the medium-educated and by about 40 log points for the low-educated, with sizeable real wage losses in the lower tail of the distribution, especially for the low-educated.

What is the nature of the estimated selection effects? Table A3.3 reports the estimated average conditional selection effect $[\sigma(X, \delta)g(\theta_\tau, Z\gamma)]$ for log wages after undoing the trans-

Table 3.3: Conditional independence tests for equality of slope coefficients in selection corrected quantile regressions after transformation (P-Values)

Covariates	Test range for τ	Medium-Educated 1995	Medium-Educated 2010	Low-Educated 1995	Low-Educated 2010
All	80-20	.385	.000	.000	.000
All	70-30	.956	1.00	.031	.394
All	60-40	.977	1.00	.541	.669
Age + Age squared	80-20	.999	.998	.474	.090
Age + Age squared	60-40	1.00	.992	.620	.751
Part-time 5years	80-20	.983	1.00	.155	.345
Part-time 5years	60-40	.993	1.00	.134	.507
Full-time 5years	80-20	1.00	.997	.035	.345
Full-time 5years	60-40	.985	.939	.672	.972
Range $(u-l)$ for Transformation (δ_u, δ_l)		80-40	75-35	80-30	75-25

Notes: P-values of specification tests under the null-hypothesis of conditional independence as in Huber and Melly (2015), testing for equality of the slope coefficients β_τ in eq. (3.6) over τ (e.g. range 80-20 denotes $\tau = .2, .25, ..., .75, .8$). Selection corrected quantile regressions based on transformed model, with tranformation based on predicted quantile difference $\sigma(X, \delta)$ in the identification-at-infinity sample, with $\delta = (\delta_u, \delta_l$ for range $u - l$. Wald tests on an equi-spaced five-percent-grid over the stated range for τ of the conditional distribution.

formation for selected values of τ and the selection probabilities $Pr(D = 1|Z) = \Phi(Z\gamma)$. These calculations disentangle the relationship between selection probabilities and the covariates X thus overstating the size of the selection effects in the sample in light of the high selection probabilities in our cases, see table 3.1. Table A3.3 covers a wide range of selection probabilities representing most of their support in the employment sample. For a very high selection probability of 99% the selection effects are zero and they grow with smaller selection probabilities. Around the median selection probabilities, the selection effects prove in the order of 10 to 20 log points across all quantiles showing sizeable positive selection into employment. Incidentally, the selection effects vary with τ, however, there is no common pattern across the four cases. For the medium-educated they tend to increase with τ, except for a very low selection probability in 1995. This suggests that for medium-educated selection effects grows with the rank in the conditional wage distribution. For the low-educated the pattern along the conditional wage distribution is less clear. The selections effects are more similar for different τ's. Specifically, for very low selection probabilities the selection effect falls with τ, similar to the medium-educated in 1995, but the selection effcts increases slighty with τ for intermediate values of the selection probabilities. While the estimated selection effects imply that there is strong positive selection into employment when selection probabilities are around 93%-97%, the range of the median in the four cases, these results do not allow us to quantify the selection affects along the unconditional wage distribution.

Based on section 3.3.5, we estimate the counterfactual distribution $T_Y(\tilde{X}, \alpha)$ to account for the different selection of observables in the total sample $\tilde{X}$, where α involves the quantile regression coefficients of log wages on X among the employed without selection correction. To account both for selection on observables and unobservables, we estimate $T_Y(\tilde{Z}, \check{\beta}, \delta, 0, \tilde{\gamma})$ setting $\theta = 0$, because there is no need for selection correction when using the full sample. Effectively, we predict wages based on the selection corrected coefficient estimates $\check{\beta}$ for sample characteristics X. Figure 3.6 displays the counterfactual wage distributions if both the unemployed and the employed were working full-time. The distribution labeled with 'Sel. on observables' accounts for the differences in observables between employed and unemployed and 'Full employment' accounts for both observables and unobservables.

There are three key similarities across the four cases. First, all counterfactual distributions lie for most part below the corresponding observed wage distributions, except for the absence of selection on observables among the medium-educated in 1995. This means that the employed in the sample are positively selected with regard to wages because the quantiles of observed wages lie above the quantiles for the entire sample including the unemployed. This reflects that wage quantiles for the unemployed would be lower than the corresponding wage quantiles for the employed. Second, the distribution accounting for observables typically lies between observed wages and the full employment distribution, implying that there is actually both positive selection on observables and on unobservables among employees. Third, the gap between observed wage quantiles and counterfactual wages is largest in the lower tail of the distribution, it falls along the distribution, and closes in the upper tail. As to be expected, the negatively selected unemployed are concentrated in the lower tail of the wage distribution.

At the same time, there are some noteworthy differences across the four cases. The figures in the upper panel of figure 3.6 shows that for the medium-educated in 1995 there is no selection on observables and strong positive selection on unobservables. The results differ for 2010, when we find small but positive selection on observables and much smaller positive selection on unobservables than in 1995. Further, the total selection effect over most of the distribution falls over time. Also for the low-educated, there are changing patterns of selection (lower panel of figure 3.6). While both types of selection seem almost equally important in 1995, the selection on observables dominates in the lower tail of the distribution and both types of selection become stronger above the median. We conclude that while selection on observables increased over time for both education groups the importance of selection on unobservables fell.

3.4.5 Keeping selection as of 1995

As last step of our analysis, we estimate counterfactual wage distributions for 2010 assuming that either selection on observables or selection on observables and unobservables had remained at its values as of 1995. Again based on section 3.3.5, we estimate the counterfactual distribution $T_Y(\tilde{X}, \alpha)$ with observables in the employment sample 1995 $\tilde{X}$ and coefficients α for wage regressions among the employed in 2010. To account both for selection on observables and unobservables, we estimate $T_Y(Z^b, \breve{\beta}^o, \delta^o, \theta^o, \tilde{\gamma}^b)$ where $\breve{\beta}^o, \delta^o, \theta^o$ represent the coefficient estimates of our selection corrected quantile regressions in $(o =)$ 2010, Z^b are the sample characteristics for the employed in 1995, and $\tilde{\gamma}^b$ the coefficients of the selection model in 1995. $Z^b\tilde{\gamma}^b$ determines the selection probabilities among the employed in 1995.

Figure 3.7 displays the two counterfactual wage distributions keeping selection as of 1995 together with the actual distribution in 2010. $T_Y(\tilde{X}, \alpha)$ is denoted as 'Observables of 1995' and $T_Y(Z^b, \breve{\beta}^o, \delta^o, \theta^o, \tilde{\gamma}^b)$ as 'Total selection of 1995'. For both education groups. the effect of the change in the selection between 1995 and 2010 is small relative to the total selection effects within both years as shown in figure 3.6. A second common finding is that the counterfactual wage distribution under the total selection as of 1995 lies below the 2010 distribution. This applies to the total range of the distribution for the medium-educated and to the range below the 70%-quantile for the low-educated. For the medium-educated, the distribution with observables as of 1995 lies between the distribution observed in 2010 and the distribution with total selection of 1995. Further, table 3.4 shows the spread of the inter-quartile gaps for the observed counterfactual wage distributions. Both counterfactual distributions displays a slightly larger wage dispersion as measured by the implied quantile differences. For the low-educated, the counterfactual with observables as of 1995 basically corresponds to the distributon of 2010, thus the change in the selection of observables does not seem to have an impact. However, the distribution under total selection of 1995 shows lower wages below the 70%-quantiles with a maximum gap around the 30%-quantile. This means that wage dispersion in the middle of the distribution, e.g. as measured by the interquartile differences, would have been higher under the selection as of 1995. However, the increase is lower when moving to the tails of the distribution.

Summing up, we conclude that with the selection of employees as of 1995 wage inequality would have been slightly higher in 2010. Despite the strong increase in wage inequality between 1995 and 2010, this findings suggests that the fall in unemployment up to 2010 by itself has not been associated with a change in the selection of employed towards higher inequality. Further, despite the fall in real wages in the lower tail of the distribution, the selection of the employed has changed towards higher wages.

Table 3.4: Counterfactual wage gaps with selection at 1995

			Low education		Medium education	
			1995	2010	1995	2010
Inter-quartile wage gap	1	Observed	0.336	0.545	0.381	0.49
		Percentage of baseyear	100	161.9	100	128.7
	2	Observables of 1995	0.336	0.501	0.381	0.472
		Percentage of baseyear	100	148.9	100	123.9
	3	Total selection of 1995	0.336	0.576	0.381	0.494
		Percentage of baseyear	100	171.11	100	129.57

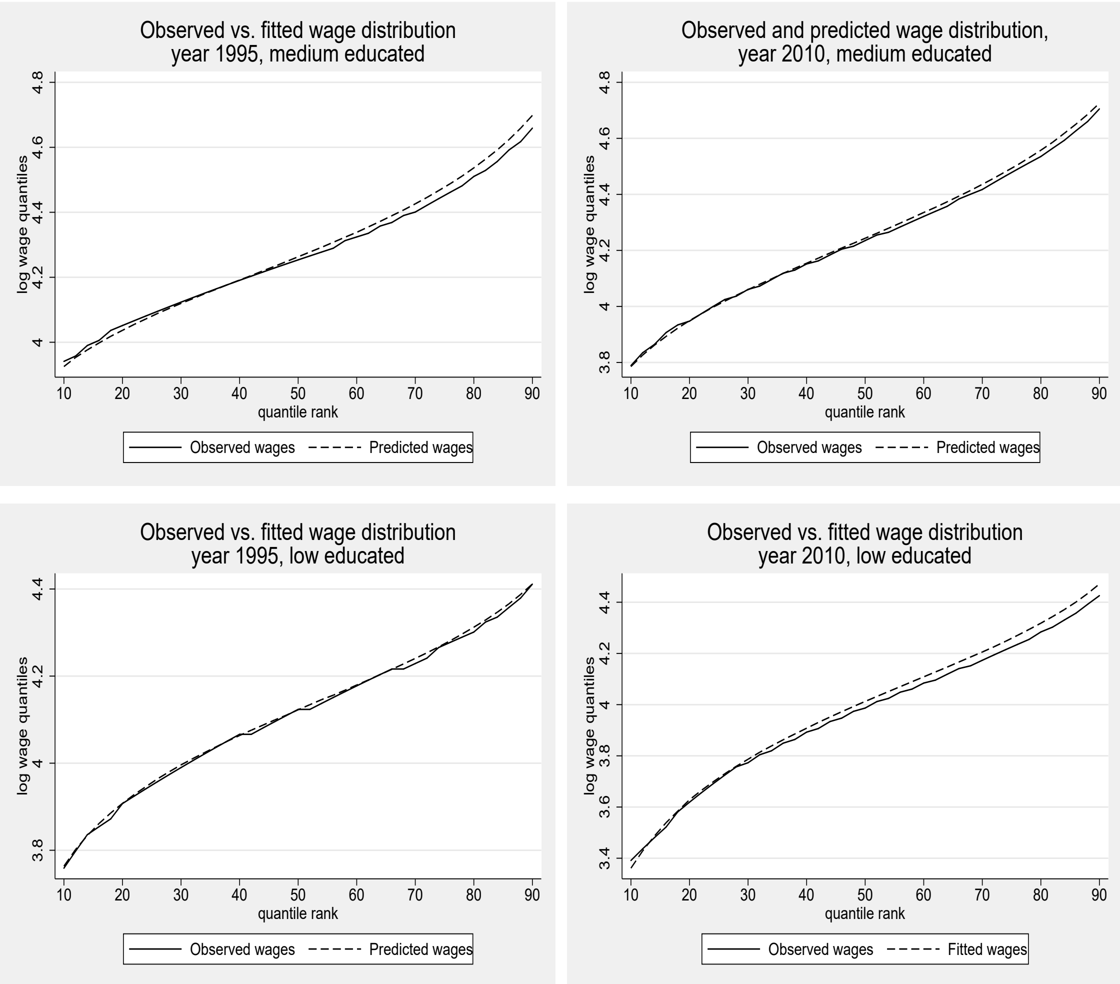
1. The observed full-time log wage distribution
2. Counterfactual full-time wage distribution which would prevail if the distribution of characteristics X was that of 1995
3. Counterfactual full-time wage distribution which would prevail if characteristics and selection with respect to unobservables was fixed at 1995

3.5 Conclusions

As its methodological contribution, this paper proposes and implements a modification of selection-corrected quantile regressions. This modification addresses Huber and Melly's (2015) concern that using a control function approach as suggested by Buchinsky (1998) is only valid under equality of the slope coefficients on the determinants of the outcome variable, which is only observed in the selected sample. We propose estimating a transformation of the outcome variable based on the identification-at-infinity assumption and then estimate selection-corrected quantile regressions for the transformed dependent variable with the goal that equality of the slope coefficient then holds. A version of the test suggested by Huber and Melly (2015) is used to guide the choice of the transformation. We emphasize that whether the transformation approach works is specific to the application. Undoing the transformation provides nonlinear selection-corrected quantile regressions for the outcome variable of interest which can be used to estimate counterfactual distributions.

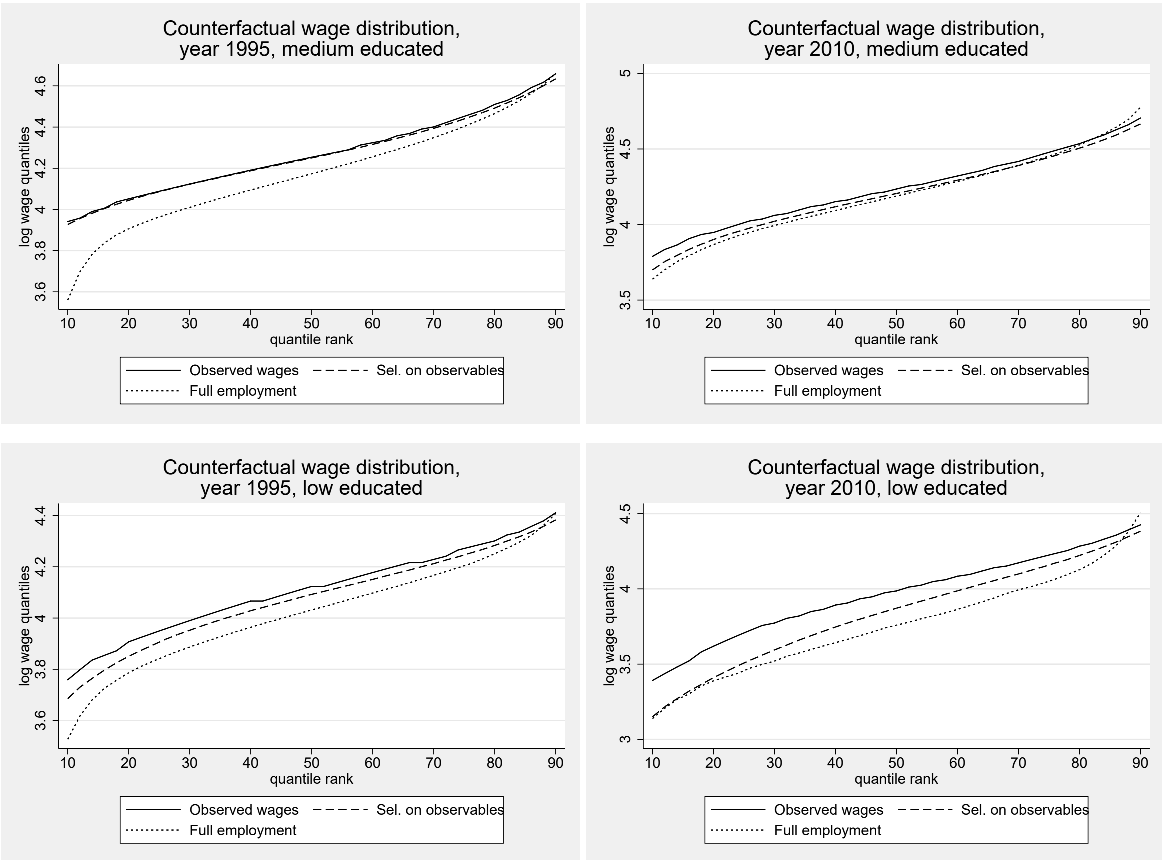
Regarding the empirical analysis of wage inequality in Germany based on the suggested modification of election-corrected quantile regressions, this papers addresses two questions. The first one is: What would the wage distribution be if all unemployed were working full time? Our analysis focuses on medium- and low-educated in the years 1995 and 2010. As to be expected, the selection of the unemployed differs strongly from the full-timers. The unemployed are negatively selected in terms of wages with respect to both observed characteristics and unobservables driving the employment probability. If the unemployed were working full-time, they would be over-represented at the bottom of the wage distribution and therefore the overall wage dispersion would be higher. Negative selection is stronger among the low-educated than it is among medium-educated workers. Our second question is: How would the wage distribution have developed if selection into full-time employment had not changed from 1995 to 2010? We find that under this counterfactual the level of wages in 2010 would have been lower in the lower and middle part of the wage distribution and wage inequality would have been slightly higher. Put differently, over time full-time workers have become less heterogeneous with regard to the factors driving wages as well as the selection into full-time work. This finding is somewhat in contrast to the existing literature emphasizing the role of composition changes in driving wage inequality (see Lemieux 2006, Dustmann et al. 2009, Biewen et al. 2018, among others). Further, selection due to unobservables did not contribute in a substantial way to the rise in within-group inequality for the medium-educated. Overall, our results suggest that the rise in wage inequality is not driven by negatively selected, previously unemployed individuals entering full-time work. A caveat to our findings is that we omit the high-educated because of the severe censoring in this group, which may explain some of the differences to the previous literature.

Figure 3.5: Actual and fitted wage distributions for employed



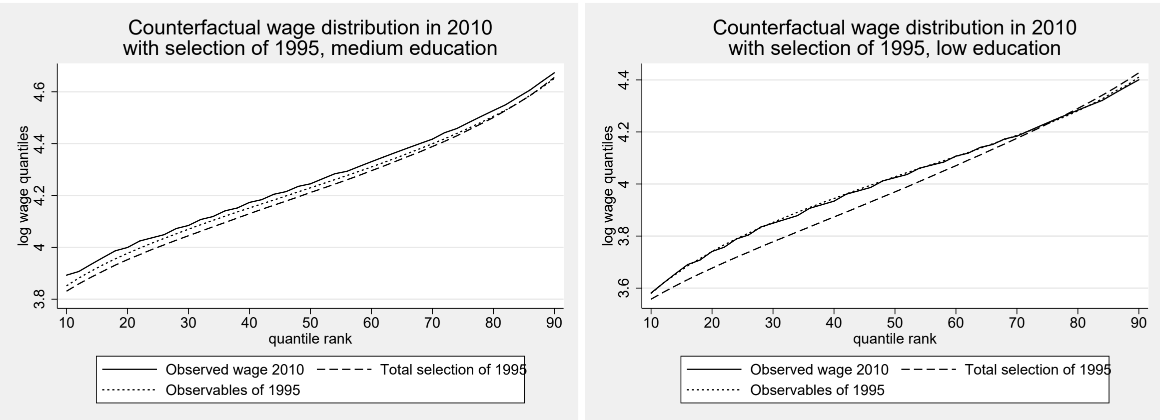
Notes: Fitted wage distribution based on Melly (2006) approach. Use model estimates for transformed data and then undo transformation. For transformed data, OLS regressions based on identification-at-infinity sample and quantile regressions with selection correction based on entire sample of employed.

Figure 3.6: Actual and counterfactual wage distributions



Notes: Counterfactual wage distributions based on Melly (2006) approach, see section 3.3.5. 'Full employment' and 'Sel.[ection] on observables' represent counterfactual wage distributions when both the unemployed and the full-time employed are working full-time. 'Sel. on observables' represents the situation where the wages are predicted based on standard quantile regressions on observed characteristics X, thus only accounting for differences in observable X. 'Full employment' represents the situation where wages are predicted based on the estimated quantile regressions with selection corrections, thus accounting both for differences in observables X and in unobservables. The counterfactual wage distributons also use predicted wages for the full-time employed.

Figure 3.7: Actual wage distribution in 2010 and counterfactual wage distribution keeping selection as of 1995



Notes: Counterfactual wage distributions based on Melly (2006) approach, see section 3.3.5. 'Full employment' and 'Sel.[ection] on observables' represent counterfactual wage distributions when both the unemployed and the full-time employed are working full-time. 'Sel. on observables' represents the situation where the wages are predicted based on standard quantile regressions on observed characteristics X, thus only accounting for differences in observable X. 'Full employment' represents the situation where wages are predicted based on the estimated quantile regressions with selection corrections, thus accounting both for differences in observables X and in unobservables. The counterfactual wage distributons also use predicted wages for the full-time employed.

3.6 Appendix

Table A3.1: Sample descriptives

	1995 Employed	1995 Unemployed	2010 Employed	2010 Unemployed
Age	38.92	38.98	41.34	39.98
Low-educated	.1325	.2738	.1196	.3609
Medium-educated	.7564	.6738	.716	.5993
High education	.1111	.0524	.1645	.0398
Days in PT last 5 years	5.703	16.61	17.18	37.5
Days in FT last 5 years	1580	885.4	1597	518.3
N	169.243	27.833	145.196	29.846
	Low-Educated			
Age	39.37	39.64	39.16	33.39
Days in PT last 5 years	4.695	32.36	29.98	10.94
Days in FT last 5 years	1444	466.3	1346	143
N	25.148	14.121	20.766	11.256
	Medium-Educated			
Age	38.59	43.37	41.43	38.06
Days in PT last 5 years	4.069	0	14.53	34.13
Days in FT last 5 years	1587	1536	1608	514.8
N	144.095	13.712	124.430	18.690

Notes: Averages of explanatory variables by subgroup. 'Days in PT/FT last 5 years' measure the number of days in part-time/full-time employment during the last five years. Source: SIAB7510, own calculations.

Table A3.2: First stage - Probit regression for full-time employment

	Medium-educ 1995	Medium-educ 2010	Low-educ 1995	Low-educ 2010
Age	-0.0468	-1.302***	0.517***	-0.559***
	(0.0556)	(0.0634)	(0.103)	(0.105)
Age squared	-0.0185***	0.139***	-0.0899***	0.0481***
	(0.00693)	(0.00788)	(0.0128)	(0.0135)
Part-time	0.193***	0.344***	0.102**	0.285***
last 5 years	(0.0239)	(0.0128)	(0.0440)	(0.0187)
Full-time	0.494***	0.625***	0.464***	0.610***
last 5 years	(0.00352)	(0.00336)	(0.00584)	(0.00556)
LS grad rate	9.791	-2.493	-	-
	(6.407)	(5.174)		
HS grad rate	29.21***	11.10**	-	-
	(6.682)	(4.904)		
Share age	3.888***	3.930***	-	-
20-30 years	(0.911)	(0.645)		
Share age	-6.386***	-5.994***	-	-
18-25 years	(2.044)	(1.484)		
Fitted prob.	-	-	1.513***	1.282***
Medium-educ			(0.442)	(0.224)
N	158,216	143,120	30,860	32,022
F_stat Instr.	20.4	29.7	11.7	32.9

Notes: Probit coefficients. Standard errors clustered at the district level in parentheses. Instruments for selection (rows 8 to 11) are measured at district level. The instruments for the medium-educated represent exogenous labor supply shocks due to variation in cohort sizes of individuals entering the labor market. For the low-educated, the instrument 'Fitted Prob. Medium' represents the average fitted employment probability for the district level based on the Probit regressions in the second and third row, respectively. F_stat Instr. denotes the F-Statistic for significance of the instruments. * p < 0:10, ** p < 0:05, *** p < 0:01

Table A3.3: Average conditional selection effect for log wages by selection probability

Medium-educated 1995					
Selection	Conditional quantile τ				
Probability	10%	25%	50%	75%	90%
99%	0.000	0.000	0.000	0.000	0.000
96%	0.094	0.136	0.175	0.199	0.198
95%	0.192	0.209	0.234	0.252	0.242
85%	0.383	0.244	0.212	0.227	0.223

Medium-educated 2010					
Selection	Conditional quantile τ				
Probability	10%	25%	50%	75%	90%
99%	0.000	0.000	0.000	0.000	0.000
97%	0.139	0.186	0.247	0.300	0.323
96%	0.258	0.257	0.278	0.301	0.303
85%	0.417	0.415	0.469	0.543	0.582

Low-educated 1995					
Selection	Conditional quantile τ				
Probability	10%	25%	50%	75%	90%
99%	0.000	0.000	0.000	0.000	0.000
95%	0.117	0.119	0.125	0.119	0.121
93%	0.228	0.239	0.243	0.236	0.240
88%	0.362	0.292	0.253	0.238	0.243

Low-educated 2010					
Selection	Conditional quantile τ				
Probability	10%	25%	50%	75%	90%
99%	0.000	0.000	0.000	0.000	0.000
96%	0.094	0.136	0.175	0.199	0.198
95%	0.192	0.209	0.234	0.252	0.242
85%	0.383	0.244	0.212	0.227	0.223

Notes: Estimated average conditional selection effect $[\sigma(X,\delta)g(\theta_\tau, Z\gamma)]$ for log wages after undoing the transformation. The selection effects are a function of the selection probability $Pr(D=1|Z) = \Phi(Z\gamma)$. Further, they differs by the τ^{th} quantile regression and by the transformation factor $\sigma X, \delta$. We calculate the average of $[\sigma(X,\delta)g(\theta_\tau, Z\gamma)]$ among all workers in the employment sample for a given selection probability, irrespective of the worker's actual selection probability. Table 3.1 reports all quartiles of the selection probabilities.

Chapter 4

Non-monotonic Selection Issues in Electoral Regression Discontinuity Designs

4.1 Introduction

The Regression Discontinuity Design (RDD), which exploits a natural experiment in order to estimate local average treatment effects (LATE), has rapidly risen in popularity among researchers in recent years (Cook 2008). Much like other causal program evaluation methods, the RDD can be biased by endogenous selection and lose internal validity. Therefore, it is of concern to researchers how cases of self-selection can be detected in advance.

In the case of the Regression Discontinuity Design, self selection invalidates the identifying assumption that the sub-populations near the assignment threshold are perfectly comparable in the absence of treatment (Lee 2008). When implementing RDDs, it is common practice in the literature to consider channels of influence trough which units of observation can influence their treatment status and perform data-driven tests of the identifying assumption (Imbens & Lemieux 2008).

We distinguish between monotonic and non-monotonic sorting dynamics. Non-monotonic sorting occurs when some individuals select into treatment while similar numbers of individuals select out of treatment. Such sorting can also happen in the form of forced de-selection by an external influence, even if all the individuals have the same treatment preference. In the literature, two kinds of tests for internal validity of the RDD are typically applied: The density based test by McCrary (2008) and checks for balanced covariate levels on both sides of the threshold. Non-monotonic sorting can not be detected with current implementations of the former and is often difficult to identify with the latter.

I will motivate the importance of finding selection issues in advance, by presenting several likely channels of influence through which individuals can manipulate their treatment assignment in an RDD. This article contributes to the literature by discussing non-monotonic selection in the RDD and developing a modified application of the McCrary specification test which can reliably detect non-monotonic sorting at the threshold. To my knowledge, the problem of non-monotonic sorting dynamics in RDD applications has not been studied in detail before. The test for non-monotonic selection works by identifying sub-samples of data whose likelihood of sorting in one direction is higher than their likelihood of sorting in the other. I then perform the density analysis on these sub-samples. If the sub-samples display an uneven density at the threshold while the full sample does not, then non-monotonic sorting is present at the threshold.

In order to illustrate the considerations for using and the workings of the specification test, I apply it to two RDD analysis, one by Lee (2008), about the incumbency advantage in United States Senate elections. By applying the test to this dataset, I follow up on the findings of Caughey & Sekohn (2011), who point out that results of close elections for US congressmen are not as randomly distributed as one would expect them to be. Also, this application illustrates that selection problems can be present even in well-established RDDs and in environments where one would, at first glance, think them unlikely. The second application is based on the first stage RDD of Dell (2015), which is also an RDD which exploits close elections.

In election settings, such as this, it is not intuitively obvious why non-monotonic selection should be an issue. The individual units of observation only have incentives to attain higher election outcomes and therefore sort themselves monotonically. However, the data contains only candidates from one party. In this case, successful monotonic sorting by each party's candidates amounts to non-monotonic sorting in the analysed sample. When applying the modified test to the election data, a suitable sub-sample which is likely to be more successful at sorting themselves above the threshold, are those candidates who's party was already in office at the time of the assignment process. The results from testing of this sub-sample indicate that a degree of sorting appears to be present. The magnitude of the estimated density discontinuity depends in part on the exact specification of the test, but the overall indications point strongly towards sorting effects. The reason why such sorting should be present is not clear cut. I arrive at the conclusion that no single factor seems to be primarily responsible, but rather a the cumulative effect of several actions with individually limited influence on election results.

The remainder of the text is organised as follows: The next section provides a quick overview of the bias introduced by endogenous selection in the Regression Discontinuity Design. It also establishes the distinction between monotonic and non-monotonic selection. This is followed by a description of the density based validity test and the modification which enables it to detect non-monotonic selection, in Section 4.3. Section 4.4 is dedicated to selection issues in the RDD application about incumbency advantages and also provides possible explanations for the testing results. Finally, conclusions resulting from the discussed properties of the test and applications are drawn.

4.2 Monotonic and non-monotonic selection in the RDD

The RDD exploits discontinuous rules, or events with discontinuous effects, to estimate local average treatment effects (LATE). Treatment is assigned according to a deterministic function, which is often a policy, law, or institutional program which assigns resources or sanctions. In addition to the outcome variable and treatment status, an independent variable is observed. This is also called the running, assignment or forcing variable. Selection into treatment is determined by a function of this variable.[1] In the Sharp Regression Discontinuity Design, assignment is completely determined by this function. In the Fuzzy Regression Discontinuity Design, the value of the running variable only partly determines participation in treatment. Since the same sorting dynamics create identification issues for both Sharp and Fuzzy RDD, this section will focus on the Sharp version of the design.

Let $X \subset \mathbb{R}$ denote the assignment variable, with $x_i \in X$ the realization of this variable for individual i and $y_i \in Y$ the outcome variable. If an individual's realization of X is above a specific threshold value c, then the individual is assigned treatment. Let $I_i(x_i) \, \epsilon \, [0, 1]$ denote treatment status. This treatment assignment mechanism implies that no overlap exists between treatment and control groups in terms of the independent variable X.

If the location of c is determined exogenously, individuals with very similar realizations of the assignment variable are likely to be similar in those characteristics which determine the outcome in the absence of treatment. In the limit, when comparing individuals directly at the threshold, the control individuals should, on average, be perfectly comparable to those receiving treatment. Identification of the LATE requires an assumption about the

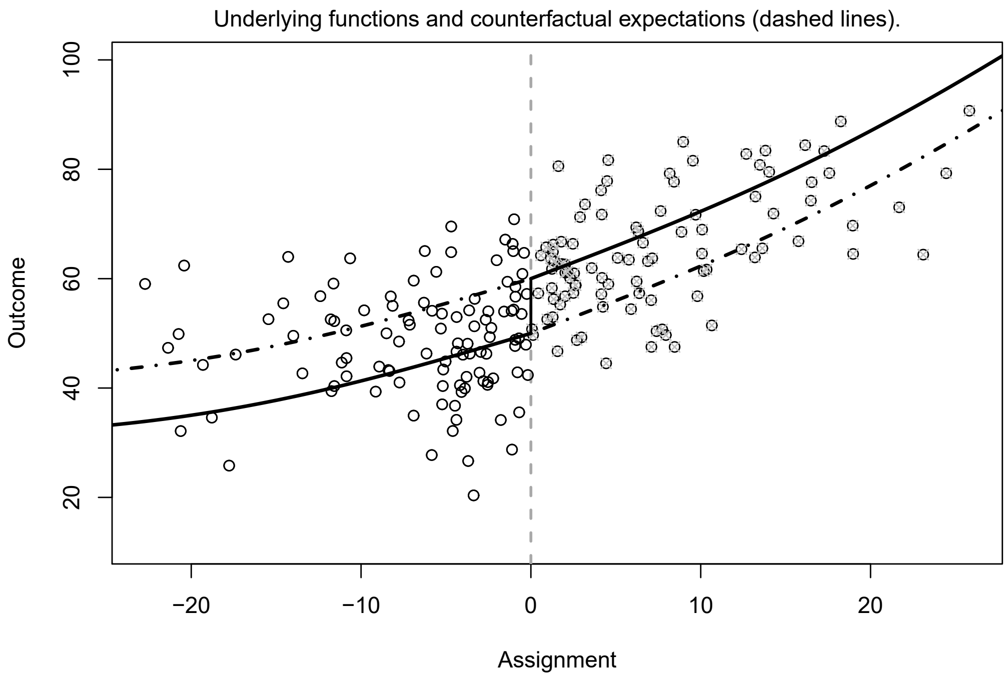
[1]For the purpose of this paper, only a single assignment variable is considered. An extension of the RDD with multiple assignment variables is discussed in Papay, Willet and Murnane (2011).

smoothness of counterfactual outcomes at the threshold: The conditional expectation functions $E[y_i(1)|x_i = c]$ for treated and $E[y_i(0)|x_i = c]$ for non-treated individuals must be continuous in c.

When this assumption holds, the LATE is identified as: $LATE = \lim_{x\downarrow c} E[y_i|x_i = x] - \lim_{x\uparrow c} E[y_i|x_i = x]$

Figure 4.1 shows an example of a fictional RDD, where the counter-factual expectations are smooth across the threshold:

Figure 4.1: Counter-factual expectations, Sharp RDD



Note: The solid lines show the true function underlying the data generating process. The dashed lines show the counter-factual expectations of the underlying function, which would prevail in the absence of treatment (lower dashed line) or if everyone was treated (upper dashed line).

The above assumption is fundamental for causal inference from RDD results. Violation of this assumption skews the RDD estimates with systematic selection bias of unknown magnitude and makes LATE estimates invalid.

In an applied setting, it might not be intuitively clear what would cause the continuity condition to be plausible. In its pure form, it is empirically not testable. To remedy this problem, Lee (2008) has linked the continuity assumption to the degree of control that individuals have over their realization of the assignment variable.

In many empirical settings, observed persons have some control over their realization of the assignment variable. They will take action to influence their realization of the assignment variable in accordance to their personal motives and underlying abilities. If individuals only roughly influence the assignment variable, then this function will include a stochastic error component. Individuals have imprecise control over X, when the density of X conditional on characteristics is continuous. This characteristic enables the empirical specification test discussed in section 4.3.

With imprecise control, treatment assignment in an area close to the threshold is "as good as randomized", meaning that the probabilities of having a value of X slightly above or below the cutoff are the same for an individual with given characteristics (Lee 2008) The continuity assumption for the potential outcomes $y_i(I_i)$ is satisfied as a consequence of random assignment near the threshold. [2]

In many applications it is assumed that all individuals have the same preference regarding treatment status. There might be a clear benefit from participation or non-participation. If individuals have uniform treatment preferences and the ability to precisely manipulate the assignment variable, then they will only shift their realization of the assignment variable in one direction.

However, this is not true for all applications. What I call non-monotonic manipulation occurs when some individuals realization of the assignment variable is shifted in one direction, while that of others is shifted towards the opposite. This can happen when the population consists of heterogeneous groups with different preferences regarding treatment assignment. A situation where this kind of manipulation was suspected was the introduction of the new German parental leave benefit (Elterngeld) on Jan. 1. 2007. The reform created incentives for some parents to postpone the birth of their child and for others to accelerate it. Birth-shifting to exploit cutoff dates is often considered unlikely, but the results of Tamm (2013) , Dickert-Conlin & Chandra (1999) and Gans & Leigh (2009) indicate that it is actively practised. This finding is important because several articles about policy evaluation use the timing of births as the cutoff for RDD analysis (Dustmann & Schönberg 2011 and Lalive 2008). In this case, Tamm (2013) finds evidence of selection into the new parental leave

[2] It is important to note that some forms of random components in the running variable are not sufficient for the continuity assumption to hold. If the random component is censored at the threshold, endogenous sorting may still be a threat to the validity of the RDD.

system, but the results for the group of parents who are expected to profit from the old system are less clear.

If both groups are of comparable size, similar numbers of individuals sort themselves to each side of the threshold. Therefore, the manipulation taking place at the threshold will not result in a jump in the density of the running variable, while still leading to systematic differences between treated and control groups.

Individual treatment preferences are not the only source of non-monotonic selection. It can also occur when realizations of the assignment variable are precisely manipulated by outside forces with contrasting preferences. In the application of Section 4.4, the sample individuals have strictly monotonic treatment preferences and some of them appear to be able to shift their assignment variable slightly above the threshold. For some other individuals however, their assignment variable is precisely manipulated to slightly below the threshold by non-sample individuals with opposing preferences. Both mechanisms can lead to problems with internal validity of the RDD, because the commonly used specification tests have trouble detecting non-monotonic manipulation.

4.3 Specification Testing

Since sorting dynamics in RDDs are often not immediately apparent, data driven specification tests are commonly used to rule out selection bias.

Two general data-driven approaches to testing RDD assumptions are available. One is to check observed covariates for smoothness at the threshold. If clear imbalances in covariate levels exist between individuals slightly below and above the threshold, the smoothness of counterfactual outcomes is unlikely. This approach relies on the availability of high-quality covariate data relevant to potential selection dynamics, which is often not available. In those cases where the characteristics affecting selection are unobserved or miss-measured, covariate tests can not rule out sorting dynamics.

A more generally applicable method for testing the identifying assumption of an RDD is the density based test developed by McCrary (2008). Applying this test to the entire sample will not detect non-monotonic sorting at the threshold. But the method used for finding discontinuous jumps in the density is also used for testing sub-samples to identify non-monotonic sorting.

The object of analysis for this test is the density function of the assignment variable. Uncensored random components in each individual's value of X imply continuity of the cumulative distribution function, conditional on underlying characteristics of each individual. And therefore imply continuity of the conditional density of the assignment variable. Continuity of the conditional density also implies continuity of the overall density of the assignment variable across the population.

If precise sorting or other types of non-random selection into treatment take place at the threshold, the density of X will not be smoothly distributed at the cutoff. It is therefore possible to check for violation of the identifying assumption by testing for continuity of the density function of the running variable at the threshold. This is done by estimating the size of a potential discontinuity in the density at the cutoff, which, in principle, is similar to performing a RDD-analysis on the density function of the assignment variable, with the treatment effect being equivalent to the deviation from continuity. Consequently, the techniques used for the density based specification test are closely related to those used in conventional RDD settings.

Under certain conditions a variation of the density testing procedure can be used to detect non-monotonic sorting. The size of the bias introduced by this kind of sorting dynamics is in direct proportion to the share of the two subgroups with contrasting treatment preferences in the sample. Within each treatment-preference group, a discontinuity in the density of observations would be present at the threshold. Therefore, testing separately for each subgroup would allow the researcher to discover these sorting dynamics.

Precise identification of the subgroups can be challenging, since the mechanics of manipulation and the individuals involved are rarely observable. If they were, data driven tests would not be required. If group membership can not be precisely determined for each individual, it is still possible to find evidence of non-monotonic manipulation. For this purpose it is sufficient to identify elements of the population for which the probability of belonging to one group is higher than that of belonging to the other group. As long as one of the sub-groups is over-represented in the tested sample and the sorting dynamics are sufficiently strong, the density test can detect those dynamics.

This approach requires additional information about the individuals compared to the straight density test, in order to determine group membership. It does however have two distinct advantages over simple tests of covariate smoothness. First, covariate data does not need to be of high quality and missing observations pose less of a problem. For example, categorical data can be used to determine subgroups suspected of self-selection. Second, this approach is helpful when self selection can not be identified in terms of a single covariate level but instead depends on interactions between covariates.

4.3.1 Details of estimation

For the specification test, a histogram of the density function is created by finely binning the running variable and assigning the frequency counts to the bin midpoints. The bins are constructed in such a way that no bin contains values of X from both sides of the threshold.

Then a Local Linear density smoother is applied separately to the histogram on each side of the threshold.[3][4] A kernel-weighted linear regression is applied to small sections of the data. Each section is defined by an evaluation point x_0 and the bandwidth h. The bin midpoints are used as regressors and the counts per bin, as regressands. [5]The kernel function that is most beneficial for RDD-applications is the triangle kernel, which shows optimal performance at boundary points.[6] Weights are assigned in a linear way, with the peak of the weight distribution at the evaluation point. At the boundary, the weight distribution is truncated and its peak lies at the boundary point itself.

A potential discontinuity in the density function will be found by performing separate regressions on both sides of and estimating the outcomes at the cutoff. The discontinuity would show up as the difference of the boundary estimates at the threshold being significantly different from zero. The specification test is then performed as a Wald-Test with the null-hypothesis that the jump in the density is zero.

It is necessary to select two tuning parameters for the estimation process: The size of the histogram bins and the bandwidth for Local Linear estimation.

The bin size has only minor effects on the results. In most applications, the estimator described above is very robust to changes in bin size, under the condition that a sufficient number of bins are covered by the bandwidth of choice.[7] I employ the bin size selection rule

---

[3]A detailed discussion of the asymptotic properties of local linear estimation can be found in Fan and Gijbels (1996). It has been shown by Hahn, Todd and Van der Klaauw (2001) that, for the purposes of the RDD, local linear estimation is highly efficient.

[4]As discussed by Lee and Card (2008), the treatment effect is asymptotically not identified for non-parametric estimation without functional form assumptions in conventional RDD applications with discrete running variables. However, this issue is not present in the density based specification test, if the running variable has continuous support. The binned running variable is not discrete in the conventional sense, because it can be defined by the researcher and the bin width can asymptotically shrink to zero when the data density approaches infinity.

[5]A detailed description of the Local Linear estimator as described in McCrary (2008) is included in the Appendix section 4.6.1.

[6]Compare Cheng et al. 1997 & Lee and Lemieux 2010 for a discussion about the merits of different kernels in the RDD.

[7]This robustness has been formally shown by McCrary (2008) and the results found in Section 4.4 are in line with those conclusions.

suggested by McCrary (2008), which is a variation of the widely used Scott's rule for bin size selection.[8]

Both more critical and more difficult is the choice of the bandwidth. It is a measure of the flexibility of the local linear model. For each evaluation point, the bandwidth determines which bins, and therefore which observations, are used for the point estimator. In RDD applications this means that the bandwidth determines how close to the threshold the data is evaluated for the discontinuity estimate.

The choice of bandwidth is essentially a trade-off between reduced bias and precision of the estimates. A small bandwidth for Local Linear estimation will result in a better approximation of the underlying function and reduce the bias, since only observations closer to the cutoff are used for estimation purposes. However, the estimate will be based on a smaller number of observations, which will reduce the precision of the result.[9]

Bandwidth choice for non-parametric estimation has been analysed in detail in the literature and a number of solutions have been proposed.[10] When ease and speed of implementation is a priority, as it is in the case of specification testing, so called 'rule of thumb' (ROT) bandwidth selectors are commonly used. A ROT bandwidth for the special case of density estimation at boundary points has been proposed by Fan, Gijbels (2006) and by McCrary (2008). Using the suggested procedure, I fit a fourth-order polynomial model to each side of the histogram and choose the bandwidth depending on the mean squared error and the curvature of the fitted model.[11] However, since the suggestions for the best bandwidth selection technique vary wildly in the literature, I treat the ROT bandwidth as a starting point and calculate tests for a wide range of bandwidths.

[8] The suggested bin size is $b = 2\hat{\sigma}N^{-\frac{1}{2}}$, with $\hat{\sigma}$ being the standard deviation of the assignment variable in the sample. See Scott (1979).

[9] As part of a discussion of the asymptotic properties of local linear estimation at boundary points, it has been shown by Hahn, Todd and Van der Klaauw (2001) that the optimal bandwidth converges to zero at a rate of $N^{-\frac{1}{5}}$, when the sample size approaches infinity. Implying that the bandwidth should be proportional to $N^{-\frac{1}{5}}$.

[10] See Pagan and Ullah (1999) for practical results from subjective bandwidth choice. Cheng (1997) and Imbens, Kalyanaraman (forthcoming) for presentations of plug-in methods. Fan and Gijbels (1996) for a "rule of thumb" for bandwidth selection. And Ludwig & Miller (2005) for a cross-validation technique.

[11] Compare Appendix 4.6.2 for a description of the ROT.

4.4 Empirical Applications

The applications analysed in this section are an RDD by Lee (2008), about the political landscape in the United States, and an RDD by Dell (2015), about the effects of the partisanship of mayors on violent crime in Mexico.

Lee determines the inherent vote share advantage which candidates for the House of Representatives receive if their party is the incumbent at the time of election. The hypothesis is that, individual characteristics being equal, those candidates whose party is in office at the time of the election have advantages over their competitors in terms of the vote share. Specification tests for the whole sample reject the presence of sorting effects. However, when applying the sub-sample test for non-monotonic sorting, the results indicate some level of precise selection at the threshold. This is an instance where the selection preferences of sample individuals are monotonic, but where the assignment variable can be subject to manipulation by exogenous agents, in this case the opposing Republican candidates.

Dell (2015) treats the average vote share of different party candidates as a proxy for socio-economic characteristics of the mayoral district. The hypothesis is then that districts in which a party barely won are, on average, comparable to those where the party barely lost. In contrast to the article by Dell, who uses election results for the Partido Accion Nacional (PAN), I consider election results for the Partido Revolucionario Institucional (PRI), which allow for higher numbers of observations. While Mexican elections are, in principle, not immune to experiencing manipulation of vote shares, results of the specification test indicate no sorting issues in this sample (BBC News, 2012). Both applications showcase why we should be aware of the unexpected ways in which non-monotonic sorting can affect RDDs.

4.4.1 Testing the full sample of the incumbency Regression Discontinuity Design

Treatment, in the form of incumbency, is assigned when the vote share difference of a party crosses the threshold at zero percent. The vote share difference is defined as the percentage difference in vote shares between a candidate and his next closest contender. This value is positive for the winner of the election and negative for the losers. Only candidates of the Democratic Party are included in the sample. The results for Republican candidates are expected to reversely mimic those of the Democrats in the majority of cases.

The assignment variable is the vote share difference at time t. This value is centred by definition, so that the threshold value c lies at zero. All districts with Democrat vote share differences to the right of the cutoff have Democrat incumbents at the time of the next election in period $t+1$. The indicator $I_{it+1} = 1[VS_{it} \geqq 0.5]$ describes the incumbency status of the candidate's party.

The outcome variable is the party vote share in the election at time period $t+1$. In the application, Democrat vote share in the following election (VS_{it+1}) is regressed on the Democrat vote share difference in the previous one and on a vector of candidate characteristics (w_{it+1}).

$$VS_{it+1} = \alpha_{t+1} w_{it+1} + \beta_{t+1} I_{it+1} + \gamma_{t+1} VS_{it} + e_{it+1} \tag{4.1}$$

with $E[e_{it+1}|w_{it+1}, VS_{it}] = 0$. The RDD is necessary because w_{it+1}, VS_{it} and I_{it+1} are all correlated with w_{it}. By performing parametric regressions separately on both sides of the cutoff, Lee (2008) finds that there is an incumbency advantage of about 7.8 percent of the vote share in the data.

The identifying condition of $f(VS_{it}|w_{it})$ being continuous in VS depends on the assumption that election results contain a substantial random component, because many factors influencing election outcomes are beyond any candidate's control. For example, weather or traffic conditions can influence election turnouts. However, some opportunities for precise manipulation in political elections have been observed and I discuss them in Section 4.4.5.

A small range of available covariates: past political experience, number of election runs, party vote share in $t-1$ and the probability of the party winning the election in $t-1$, show balanced levels within a 5 percent margin of the cutoff.

When applying the density based specification check described in Section 4.3 to the data, no significant sorting can be detected.[12][13] To establish the robustness of the results for different values of the tuning parameters, I performed the test with the reference bin size and bandwidth, as well as fractions and multiples of both reference values.

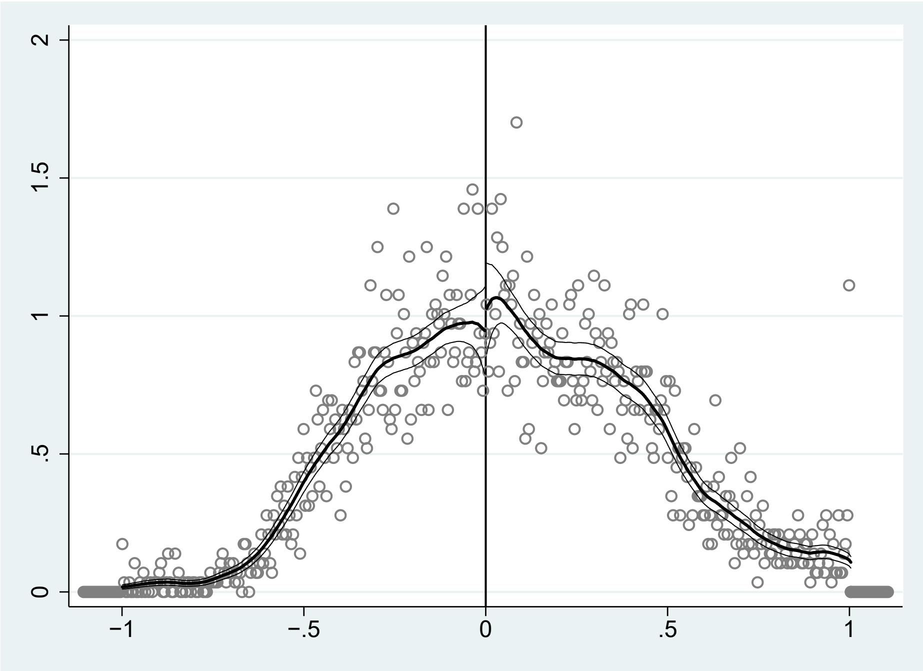
Table 4.1 shows the discontinuity estimates for all combinations of tuning parameters and Figure 4.2 shows the fitted model. The version in this graph provides, upon visual inspection, the best approximation of the data close to the cutoff of all tested variations.

---

[12] A dataset containing the information for the Lee study has been obtained from the Mostly Harmless Econometrics Data Archive: `http://economics.mit.edu/faculty/angrist/data1/mhe` (last visited 15.05.2014).

[13] The sample is trimmed at the extreme ends of the forcing variable to remove outliers and improve the clarity of plots without affecting local linear estimators.

Figure 4.2: Density estimates for Democratic candidates



Note: Dark lines show flexible local linear kernel regressions fitted to binned observations, separately on both sides of the threshold. 95% confidence bands in thin lines. Bin size of 0.0048, bandwidth of 0.1061.

The results indicate a very smooth distribution of election results at the threshold. No discontinuity estimate exceeds two standard deviations and the estimated differences in log-densities at the threshold range between 0.0104 and a maximum of 0.1831. For the entire range of bandwidths and bin sizes, t-tests do not reject the null hypothesis of a smooth distribution. On the aggregate level, the density function is continuous at the threshold.

4.4.2 Testing the sub-sample of incumbent Democratic candidates

In a recent article, Caughey and Sekhon (2011) have questioned whether the outcome of close elections to the U.S. House really is as randomised as Lee (2008) assumes. They show that a number of relevant covariates are not well balanced at the threshold.[14] Covariate imbalance is greatest away from the threshold, diminishes when looking at observations closer to the threshold, and increases again for extremely close elections. They report that covariates become more balanced within shrinking margins around the threshold, down to a margin of five percent. This finding is in line with the results from Lee (2008). However, for

[14]These covariates include, among others, the political experience advantages for Republican and Democratic candidates, campaign money spent and donation funds received.

Table 4.1: Estimated discontinuities in the density at the threshold, test results for the full sample of Democratic candidates

	Quarter the reference bandwidth 0.0531	Half the reference bandwidth 0.1061	Reference bandwidth 0.2123	Twice the reference bandwidth 0.4245	Four times the reference bandwidth 0.8491
Reference bin size 0.0097	0.1359 [0.1871] *0.4676*	0.0844 [0.1247] *0.4985*	0.1258 [0.0856] *0.1417*	0.0484 [0.0605] *0.4237*	-0.0108 [0.0414] *0.7942*
Half the reference bin size 0.0048	0.1004 [0.1851] *0.5875*	0.0800 [0.1245] *0.5205*	0.1163 [0.0856] *0.1743*	0.0454 [0.0605] *0.4530*	-0.0116 [0.0414] *0.7793*
Twice the reference bin size 0.0194	0.1831 [0.1827] *0.3163*	0.0968 [0.1247] *0.4376*	0.1304 [0.0856] *0.1277*	0.0484 [0.0604] *0.4229*	-0.0104 [0.0414] *0.8017*

Note: The table shows the size of the estimated discontinuity in the density at the threshold, with standard errors in brackets and p-values in italics. The reference bandwidth and bin size are selected according to the ROT approach (compare appendix section 4.6.2).

smaller margins, especially those of less than one percent, covariates become less balanced.

As causes for this behaviour, monotonic manipulation of the running variable is ruled out by the density based test. Non-monotonic sorting issues in the sense that equal numbers of individuals with opposing treatment preferences sort themselves to each side is also not possible, because the sample of Democratic candidates have strict monotonic treatment preferences.

An explanation could be a combination of monotonic manipulation by sample individuals and external forces with different treatment preferences: From the perspective of each party, manipulation is always strictly monotonic positive, since winning the election is the primary goal of any candidate running for office. However, the aggregate situation is not as clear, since both Democrat and Republican candidates engage in manipulative activities.[15] If a Democrat party candidate was able to precisely control his vote share, he would realize

[15] For this argument, a strict two-party system is assumed. This assumption closely but not entirely reflects the political realities of post-war elections to the U.S. House of Representatives.

a vote share margin of victory marginally above the threshold. As a direct consequence, his Republican contender would receive a vote share slightly below that of the Democrat candidate, and therefore marginally below the threshold.

If some candidates from one party have the ability and opportunity to manipulate their vote shares, we have to assume that the other party would possess the same capabilities. Consequently, a number of Republican candidates would also be able to precisely manipulate their vote share. Those candidates would win a disproportionate number of close elections, causing a similar number of Democrat candidates to barely lose the elections. This would lead to a discontinuous jump downwards in the density of Democrat vote shares.

If comparable amounts of successful precise manipulation were achieved by both Democrats and Republicans, the effects would mask each other over the entire sample and make detection by means of the density test impossible.

We can not identify in which elections which candidate might have successfully engaged in precise sorting. However, it is enough to identify sub-samples of candidates who have an above-average probability of precisely manipulating their assignment variable or having it manipulated by the opposing candidate.

One such sub-group would be those candidates whose party already was the incumbent party at the time of the election which determines assignment. This is in line with the finding that the covariate imbalances in close elections found by Caughey and Sekhon (2011) are especially pronounced between candidates running for the incumbent party and the candidates of the challenging party. The incumbent party is more deeply interwoven with the administrative institutions and therefore has potentially greater influence on the election process.

Another possible subgroup with a higher chance of successful manipulation would be those candidates who's party holds the office of secretary of state, who is in charge of the elections, or who's party provides the state governor.[16]

One might ask if the increase in the probability of winning of incumbent party candidates is not just the expected effect of the incumbency advantage from the previous election. Indeed, when looking at the aggregate of all incumbent party candidates, they have substantially higher chances of winning the next election. However, under the identifying assumption of the RDD, this should not be true for close elections. Instead, incumbent party candi-

[16]In the case of the state governor, I could not detect similar evidence of sorting mechanisms.

dates should be winning more often with higher margins of victory, since the incumbency advantage is reflected in a higher vote share.

If the incumbent party candidates are more successful at shifting their vote share precisely upwards, the mechanics leading to a discontinuity in the subgroup density would work as follows:

I take Equation 4.1 as starting point, with the vote share at $t+1$ as the outcome of interest for the RDD. For the specification test, the density of the democrat vote share at time t is analysed. And $t-1$ is the election which determines incumbency status for the purpose of subgroup testing. The density of the winners of the election in $t-1$, who will be the incumbents in time period t, $f(VS_{it-1}|I_{it-1}=1)$, is truncated at zero (compare Figure 4.3).

If no selection process is at work, then the distribution of election results in the next period, t, will appear like that of Figure 4.4. The results for incumbent party candidates, the winning party of the election in $t-1$, are concentrated at the upper end, because β_t, the vote share advantage from incumbency, shifts them upwards. The model determining vote share for this election follows the same concept as Equation 4.1:

$$VS_{it} = \alpha_t w_{it} + \beta_t I_{it} + \gamma_t VS_{it-1} + e_{it} \tag{4.2}$$

Under the no-sorting assumption, since individual characteristics w are continuously distributed, $f(VS_{it}|w_{it})$ is continuous in VS. The density of VS_t is smooth across the threshold for all groups of candidates. If however the non-monotonic sorting dynamics described earlier are present, then we would expect densities like to those in Figure 4.5. Incumbent candidates of both parties have a higher chance of winning close elections. For our sample of Democrat candidates, this leads to discontinuous jumps in the density of both the winning and losing party candidates of the previous election. For the winning ones, the discontinuity of value $\delta_1 > 0$ is caused by their ability to influence close elections in their favour. For the losing party candidates, the discontinuity of value $\delta_2 < 0$ is caused by their opponents ability to win close elections. Over the density of the entire sample of Democrat candidates, a discontinuity of size $\delta = \delta_1 + \delta_2$ is present. When both parties are very similar in terms of average political influence over time, both discontinuities cancel out and the density for the full sample does not show a gap at the threshold.

4.4.3 Results of sub-group testing

These results from the test on the sub-sample of incumbents at time t differ sharply from the ones for the aggregate sample. Estimated discontinuities in the density at the cutoff are considerably larger for all bandwidths and bin sizes. The estimates vary in size depending on the choice of tuning parameters, generally exceed two standard deviations and are always larger than one standard deviation. Plotting the local linear smoother over the histogram in Figures 4.6 and 4.7 shows a sharp downturn in the chance of barely loosing an election for the sub-sample. This indicates that of the incumbent Democrat candidates, the vast majority wins the elections they are running in, strengthening the notion that incumbent party candidates on average possess superior means of securing election wins. Especially when restricting the analysis to very close elections, by selecting a bandwidth below one percent of the vote share difference, the average chance of winning the election is significantly higher if the candidate's own party is in power.

Table 4.2: Estimated discontinuities in the density at the threshold, test results for the sub-sample of Incumbent-party candidates

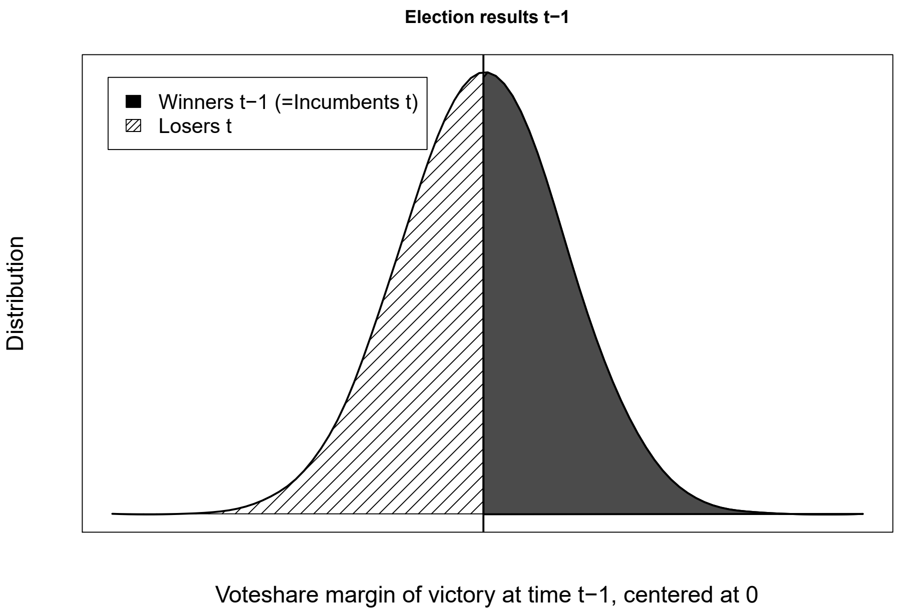
	Quarter reference bandwidth	Half reference bandwidth	Reference bandwidth	Double reference bandwidth
Discontinuity	0.882	0.344	0.310	0.542
Standard Error	0.356	0.216	0.143	0.103
P-value	0.013	0.111	0.030	0.000
Bandwidth	0.049	0.098	0.196	0.391

Note: The first row shows the size of the estimated discontinuity in the density at the threshold. The reference bandwidth and bin size are selected according to the ROT approach (compare appendix section 4.6.2)

Some variation is visible in the results, depending on choice of the bandwidth. Because of this sensitivity, I performed the test for a finely gridded range of bandwidths ranging from 0.02 to 0.25, maintaining the reference bin size of 0.0098 (Figure 4.8). As would be expected, precision of the estimates degrades with shrinking bandwidths, due to lower observation counts available within the bandwidth. For bandwidths larger than of 2% of the vote share, as well as for bandwidths smaller than one percent, significant discontinuities are estimated. As with previous applications, the results are relatively stable under variations in bin size. The null hypothesis of continuity of the vote share difference is not rejected for half the reference bandwidth, even though it is rejected at all other bandwidths. Figure

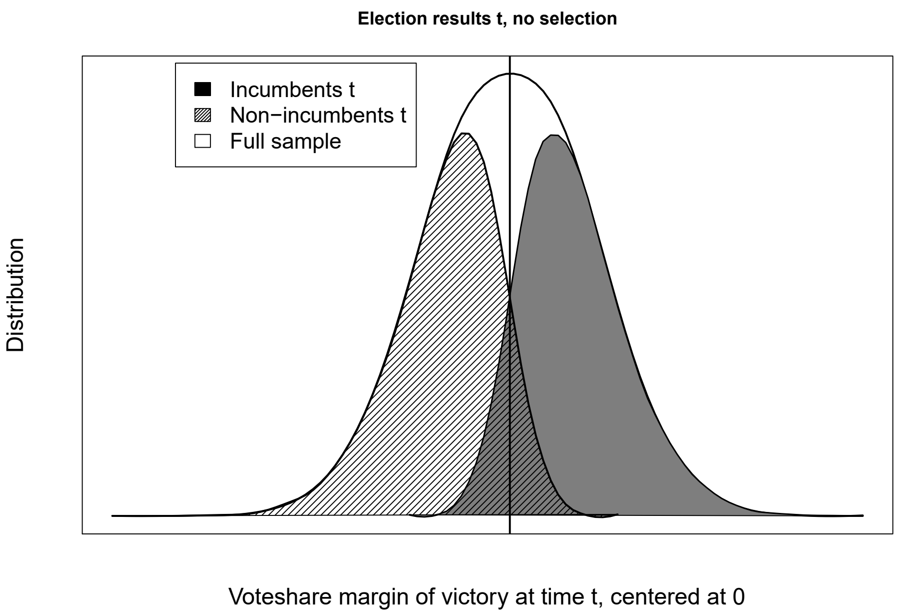
4.8 shows a more detailed view of this phenomenon. Tests for the discontinuity being non-zero are significant at the 5% level for all bandwidths smaller then 0.08 and larger then 0.2. When considering 10% significance levels, the null is rejected everywhere except for a small range of bandwidths between 0.11 and 0.125. The graph of the significance level exhibits a hump in the area of the halved reference bandwidth. While the discontinuities are not always significant at very high levels for all bandwidth choices, a sharp increase in the differences between treated and control candidates at the threshold, compared to the full sample analysis, can not be denied. This strongly hints at substantial differences in the behaviour at the boundary between incumbent party candidates and candidates of the challenging party. Even more important, the estimated jump in the density at the cutoff actually increases for very small bandwidths, when only data from the closest elections is used. This result is in line with the findings of Caughey and Sekhon (2011), who report that differences in covariate values increase for extremely close elections with vote share differences of one percent or lower, after having converged before with shrinking margins.

Figure 4.3: Determination of incumbent status in t-1



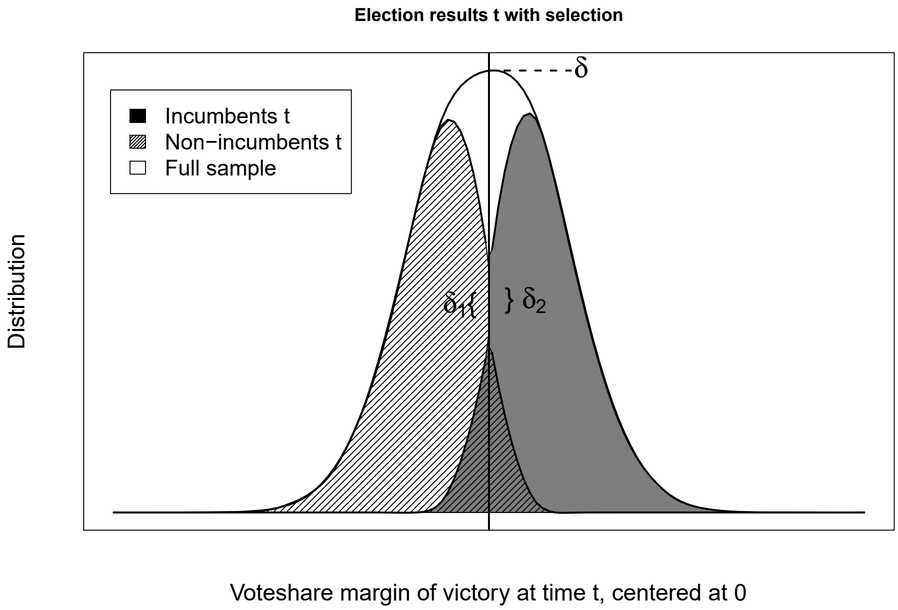
Note: Winners in $t-1$ (=incumbents in t) and losers in $t-1$ (=non-incumbents in t) are determined by vote share and precisely separated at the threshold.

Figure 4.4: Assumed distribution at t without sorting



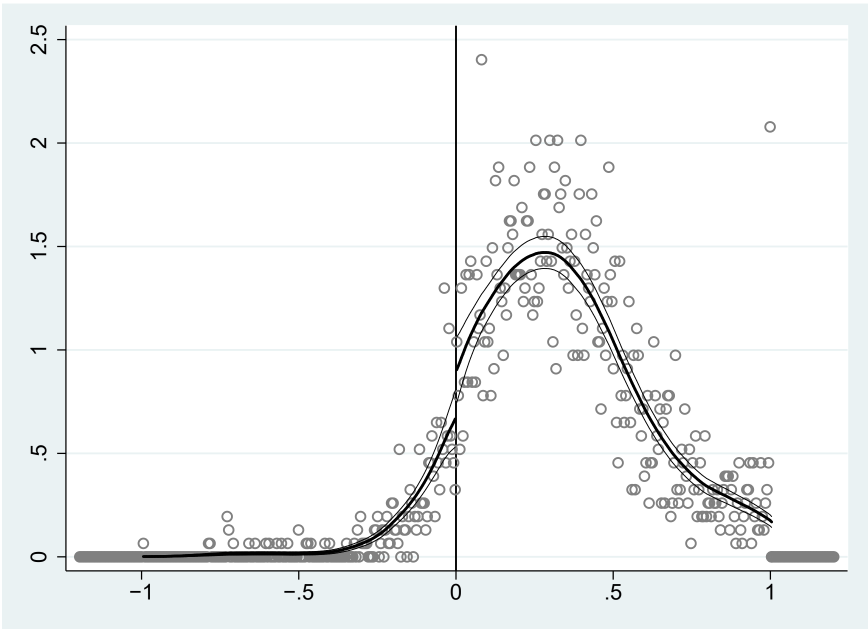
Note: Assuming no sorting at the threshold, incumbents in t (=winners of $t-1$) have a higher overall chance of winning, but not in close elections. The reverse is true for non-incumbents. Therefore the densities of both groups are smooth across the threshold and so is the density of the full sample.

Figure 4.5: Distribution at t with sorting



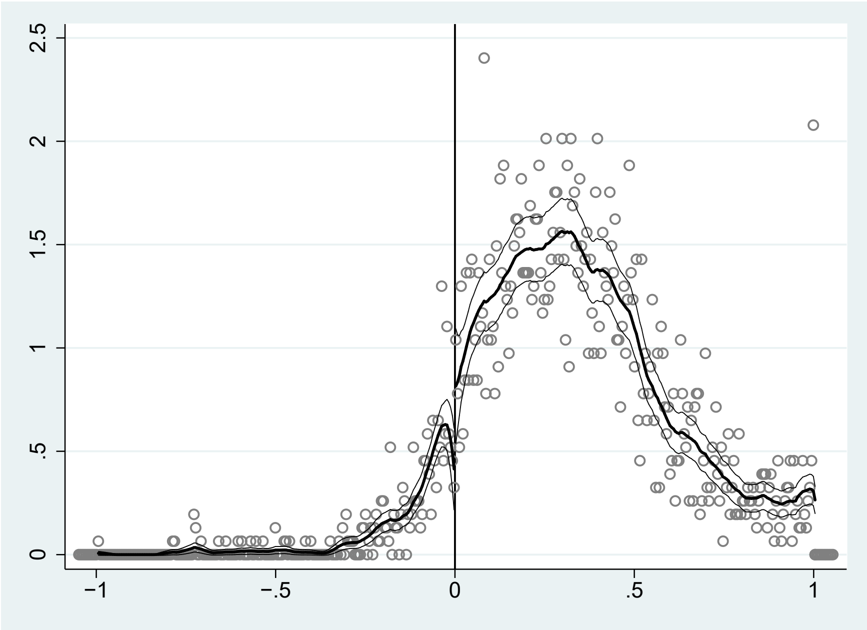
Note: If incumbents (of both parties) sort above the threshold, the densities of both incumbents and non-incumbents will show discontinuous jumps of similar magnitude but opposite sign at the threshold. Therefore the density of the full sample remains smooth across the threshold.

Figure 4.6: Density estimates for incumbent Democratic candidates



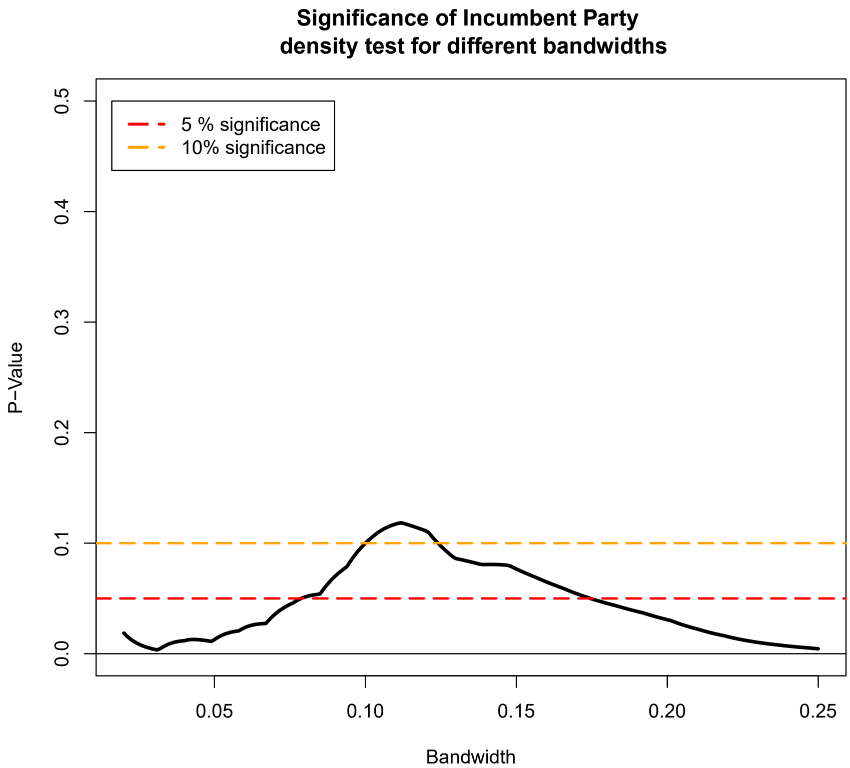
Note: Dark lines show flexible local linear kernel regressions fitted to binned observations, separately on both sides of the threshold. 95% confidence bands in thin lines. Bin size of 0.0049, bandwidth of 0.2014.

Figure 4.7: Density estimates for incumbent Democratic candidates



Note: Dark lines show flexible local linear kernel regressions fitted to binned observations, separately on both sides of the threshold. 95% confidence bands in thin lines. Bin size of 0.0049, bandwidth of 0.0504.

Figure 4.8: Significance levels depending on bandwidth



Note: Only for a small range of bandwidths for the local linear kernel regression can we confidently reject the hypothesis of no discontinuity in the density of incumbent party candidates at the threshold.

4.4.4 Application to Mexican mayoral elections

In Dell (2015), the author uses an RDD to analyse the causal effect of mayor partisanship on drug related homicides in Mexico from 2007 to 2010. When conservative president Felipe Calderon came to power in 2006 his party, the Partido Acción Nacional (PAN) spearheaded the “war on drugs”. Municipalities which were won by PAN mayors experienced more frequent and effective police activity against drug trafficking organisations. Dell’s RDD strategy is based on the concept that, on average, municipalities in which the PAN barely won the mayor’s office are comparable with municipalities in which it barely lost. Exploiting this natural experiment, the article shows an average difference in drug related homicides of 33 per 100.000 inhabitants at the vote share threshold.

As with the application by Lee (2008), sorting dynamics can not be ruled out a priori in this electoral RDD. Surrounding characteristics of the elections are different, with mayoral elections during the observed time frame not being balanced between two parties. Mexico has three prominent parties, with the PAN, the Partido Revolucionario Institucional (PRI) and the Party of the Democratic Revolution (PRD) alternating in strength across municipalities. Within the sample, the PRI wins 59% of the elections and the PAN 24%. Therefore the mechanics of the relatively strict two party system in the Lee example are no longer present. Since the three parties are, on average, not equal in political strength, sorting into treatment is less likely to be masked by equal magnitudes of sorting out of treatment, if non-monotonic selection takes place. The full sample density test has higher chances of detecting sorting behaviour.

The replication files provided in the online appendix of Dell (2015) are limited to elections within a +5% and -5% vote share interval around the threshold. I perform the density based test analogous to Sections 4.4.1 and 4.4.2 in first for the the full sample of PAN candidates, with the ROT selected bandwidth and a range of tuning parameters spanning 2% to 5% of the vote share.

The results in Table 4.3 show no indication of selection issues. Fitting a linear model on the 5% vote share bandwidth, the estimated discontinuity is almost zero. At all bandwidths, the discontinuity is not significant. Although it is not small at the reference bandwidth the confidence bands increase with smaller bandwidths and therefore diminishing numbers of observations. If the PAN was capable of deciding significantly more or less close elections for itself than the other two parties combined, it would show up as a discontinuity.

When restricting the sample to those candidates in whose municipalities the PAN was

Table 4.3: Density test results for all PAN candidates

	Refrence Band-width	Bandwidth 3 % vote share	Bandwidth 4 % vote share	Bandwidth 5 % vote share
Discontinuity	-0.412	-0.221	-0.133	-0.039
Standard Error	0.314	0.246	0.218	0.194
P-value	0.190	0.368	0.542	0.840
Bandwidth	0.020	0.030	0.040	0.050

Note: The first row shows the size of the estimated discontinuity in the density at the threshold. The reference bandwidth and bin size are selected according to the ROT approach (compare appendix section 4.6.2)

Table 4.4: Density test results for for PAN incumbent candidate sub-sample

	Reference Band-width	Bandwidth 3 % vote share	Bandwidth 4 % vote share	Bandwidth 5 % vote share
Discontinuity	-1.953	-0.711	-0.402	-0.276
Standard Error	1.395	0.600	0.469	0.401
P-value	0.161	0.236	0.392	0.491
Bandwidth	0.017	0.030	0.040	0.050

Note: The first row shows the size of the estimated discontinuity in the density at the threshold. The reference bandwidth and bin size are selected according to the ROT approach (compare appendix section 4.6.2)

already in office at the time of election, analogous to the incumbency sub-sample of Section 4.4.2, Table 4.4 shows a similar picture to the results in Table 4.3. For all bandwidths, the discontinuity is not significant at any level, but it is larger across the board, compared with the full sample estimates. These results reinforce the testing done by Dell (2015), who reports no hints of sorting behaviour in Mexican mayoral elections. Even though the potential for precise manipulation of the vote share is not not lower in Mexican mayoral elections,

4.4.5 Discussion

The results from the sub-sample specification test performed in section 4.4.3 suggest the presence of non-monotonic sorting issues in this particular RDD application. Such a result poses the question why it should be possible for incumbent party candidates to influence close elections in a way which systematically increases their chances of victory. A number of possible channels of influence exist which might allow for relatively precise manipulation of election results. The first is precise influence on the vote count by non-democratic means. Either in the form of ex-ante activities, such as the buying of votes, or in the form of ex-post manipulation, an example of which would be miss-reporting of vote counts. While the literature reports no evidence that vote fraud would be a regular or systematic issue for elections in western democracies, the possibility can not be ruled out completely (Alvarez and Hall 2006). Events like the recount of the Florida presidential elections votes in 2000 occasionally make the headlines and spark scepticism about the validity of election processes ((Lott Jr., 2001)). Especially in the case of very close elections, where only a relatively small amount of manipulation would be necessary to turn the results. As vote count manipulation is, by nature, a clandestine activity, small scale manipulation could potentially go undetected in the majority of cases. Occasional incidents of verified vote rigging show that democratic safeguards are not always effective. One such incident of electoral fraud in the U.S. was the Texas senatorial runoff election in 1948. The election was a very close one, with both campaign offices being aware of that fact. Caro (1990) reports that the campaign staff of challenger Lyndon Johnson influenced voters by directly paying out cash, appointing sympathetic election officials and bribing influential local bosses who would send their employees and dependants to vote for Johnson. Additionally, allies of Johnson later confirmed acts of ex-post manipulation, where election officials would interfere with the counting and tabulation of votes, to ensure favourable results for their party.[17] In the end, Johnson won the election by a very small margin of only 87 votes out of about one million votes total (Caro 1990).

Yet, even though media attention devoted to elections, especially close ones, has increased during the last decades, the number of confirmed incidents of vote rigging has been small and ever declining.[18] If electoral fraud was common in U.S elections, the increase in media coverage and scrutiny should have led to an increase in contested elections. But the fraction of contested elections for the U.S. House has constantly and substantially decreased during the postwar period, as Jenkins (2004) reports. Even though some cases of illegal manipulation of the voting process surely have not been discovered, it is likely that the fraction of elections which were decided by fraudulent actions, is quite small. These findings suggest that vote fraud has rarely been a deciding factor for elections in western democracies, during the time period covered by the data. In all likelihood, vote fraud alone is not sufficient for

[17]Election judge Luis Salas, who was involved in the tabulating of votes, later admitted the fraudulent manipulation of election results. See Caro (1990).

[18]The work of Campbell (2005) only reports a minimal number of elections where fraudulent activities were discovered over the last decades.

explaining the dynamics at the threshold apparent in the Lee data.

In the case of Mexican mayoral elections, we do not observe precise sorting behaviour which might result from electoral fraud, even though elections in Mexico are sometimes subject to fraudulent behaviour, as reported by McCann (1998) and Lehoucq (2003). However, the stakes in mayoral elections are not as high as those for representatives on the national level, which might reduce incentives for fraud. Another possible explanation would be that on the local levels the consequences for fraud by the incumbent are so weak, that vote-rigging results in clear wins or losses, instead of close elections.

A second mechanism, which might play an important role in deciding close elections, is the use of 'emergency'-resources. Political parties will allocate more resources to close elections, than they would to those where they expect a clear win or loss. In these elections, the marginal effect of resources spent is greatest. Even activities which are extremely costly can be deemed worthwhile, if only a minimal shift of the vote count is necessary to win the election. These resources are not necessarily of a monetary nature, but can also take the form of organisational capacities or the ability for dealing with unforeseen events. Particularly, parties and candidates in extremely close elections will perform actions which are costly in terms of political influence or long term credibility, in order to win the race. Examples of such actions would be the trading of political favours, populist promises or policies which are not in line with the party platform. They might also make use of one-time resources, like calling in favours from influential groups or individuals. It is hardly possibly to measure a candidate's emergency-resources, which prevents researchers from analysing potential imbalances in this covariate. Candidates with superior financial and organisational resources are better informed about the current state of the race and can react more quickly and effectively to problems which their supporters might encounter. Those actions combined would exacerbate existing imbalances in terms of campaign funds for very close elections. They would also lead to a situation where the candidate with access to superior 'emergency'-resources has a distinct advantage. The actual magnitude of this specific kind of resource is likely unobservable. However, it is reasonable to assume that incumbent party candidates usually do possess an advantage in that regard.

For the Mexican mayoral elections, as noted earlier, the stakes are not as high and the resources available to the candidates are orders of magnitude below those for US House representatives. It is therefore likely that candidates are unable to monitor ongoing elections as quickly and effectively, and might lack the necessary information and resources to precisely influence elections in progress. This, in turn, would lead to a larger random component in the vote share outcome, which strengthens the RDD identification strategy.

The third, channel by which close elections can be non-randomly decided is the legal influence which incumbent parties have over the political administration. Among those measures are voter-suppression tactics like restrictive voter ID laws and targeted placement of polling stations.[19] For example by increasing the density and convenience of the polling infrastructure in areas with historically strong support for the incumbent party. Selective recounting of votes is another instrument which may allow for relatively precise manipulation of the vote score. Increased influence allows a party to lobby more effectively for or against vote recounts in districts which they expect to favour their, or the opposing candidate, respectively. Nevertheless, the number of times where recounts have reversed the results in a U.S. House race has been very small. According to Caughey and Sekhon (2011), vote recounts only had a pivotal effect on the election results in less than ten percent of the sample elections in which recounts did happen. While the result was reversed in favour of the incumbent party candidate in all reported cases, the low overall percentage of pivotal recounts rules out recounting as the main factor in explaining the observed imbalances. Incumbent parties do have other means, by which they could influence the vote share in close elections. Election officials in local offices usually do have a certain amount of discretion when dealing with unclear or provisional ballots, as is analysed by Kimball et. al. (2006). The party which has endeared itself to the administrative personnel during their last term in office will have gained an advantage as a result. Also, the partisanship of election officials can play a role in circumventing adverse conditions for the own party's supporters. One example of such practices, reported by Hauser and Holusha (2006), is that officials and judges can extend voting hours in districts which favour the party they are affiliated with. This kind of manipulation is more likely in very close elections, because the marginal effect is larger.

The channels for manipulation presented here probably do not represent all avenues by which candidates can precisely sort themselves around the threshold. While no singular main reason why sorting at the the vote percentage cutoff should be possible in U.S. House elections is apparent, it is likely that the described effects, possibly combined with undiscovered factors, lead to the observed sorting behaviour.

[19] For a discussion of the impact of voter ID laws on election outcomes in the US, see Weiser et al. (2005).

4.5 Conclusions

In this article, the issues associated with non monotonic endogenous sorting in the context of the RDD are presented and a testing method for the validity of the design is discussed. The merits of thoroughly checking the data for evidence of precise manipulation of the assignment variable are motivated by a description of the various ways by which individual units can influence their realization of the assignment variable. Such manipulation can be detected by a specification test which is designed to find discontinuities at the threshold in the density function of the running variable. When non-monotonic sorting is happening at the threshold, the testing procedure described in the literature can be modified to suit the challenge. By testing sub-samples of the data which contain disproportionate numbers of individuals who manipulate their assignment score in a single direction, non-monotonic sorting can be detected. While it is possible to detect this problem by means of covariate distributions, the density based test expands the arsenal of researchers with a method far less demanding of the quality and spectrum of covariate data.

As example applications, the well established RDD analysis of the US-House incumbency advantage by D. Lee (2008), and the Mexican mayoral election RDD by Dell (2015), are examined with regards to a special form of non-monotonic selection effects. For the US House elections, testing on the aggregate level of all Democrat candidates does not reject the hypothesis of continuity of the assignment variable at the threshold, which is in line with the results of McCrary (2008) and the specification testing performed by Lee himself. However, the situation is not quite as clear when considering non-monotonic selection. Within a sample of all Democrat-candidates, the strictly positive self-selection of Democrats would be masked by the strictly positive self-selection of Republican candidates, considering the predominantly two-party system. The sub-sample of democratic-party candidates whose party was the incumbent at $t-1$ displays unexpected behaviour of the density function at the threshold. This is the sub-sample of individuals who are most likely able to precisely influence their assignment variable, according to Caughey & Sekohn (2011). Results for this sub-sample reject the hypothesis of a smoothly distributed assignment variable at the threshold for a substantial range of tuning parameters, with significance actually increasing for very close elections. Therefore it appears likely that certain candidates possess the ability to precisely sort themselves to one side of the threshold. In the case of the Dell (2015) application, no evidence for non-monotonic sorting can be found.

So far, no comprehensive explanation is available which would explain why precise manipulation of the vote share difference should be possible in US House elections, although a number of factors which might contribute to the sorting dynamics were discussed. Qualitative analysis of candidate behaviour in close elections could shed more light on this issue. These results might lead to a reinterpretation of the incumbency advantage estimated by Lee (2008), if higher probabilities of winning close elections are common perk of being the incumbent. Since manipulation of the assignment variable is performed primarily by incumbent party candidates, it is worth considering to what extent the ability for manipulation in one election translates into a vote share advantage in the next election. From this perspective, it is worth considering if the discovered imbalances between close winners and losers introduce lead to a cumulative incumbent party advantage over multiple elections. In the latter case, sorting behaviour in subsequent elections might be an integral part of this advantage.

4.6 Appendix

4.6.1 Description of the Local Linear density smoother

Construct J bins with $j = 1...J$ and bin width b. Let J_l and J_r denote the number of bins to the left and right of the cutoff c, respectively. The bins are defined as intervals:

$$(d_j,\, d_{j+1}] \quad with \quad d_j = c - b(1 - j + J_l)$$

and bin midpoints X_j with $|X_j - d_j| = |X_j - d_{j+1}| = \frac{b}{2}$. Calculate the normalized observation counts per bin:

$$N_j = \frac{1}{Nb} \sum_{1}^{N} 1(d_j < x_i \leq d_{j+1}) \tag{4.3}$$

The histogram is then established by plotting the frequency counts $N_j$on the bin midpoints X_j.

The Local Linear estimator for a given bandwidth h and a kernel weighting function K, at $x_i = x_0$is described by:

$$\hat{y}(x_0) = \hat{\beta}_0(x_0) + \hat{\beta}_1(x_0)(x_0 - x_0)$$

with $\hat{\beta}_0(x_0)$ and $\hat{\beta}_1(x_0)$ minimizing the loss function:

$$\begin{aligned} & L\left(\hat{\beta}_0(x_0), \hat{\beta}_1(x_0)\right) \\ = & \sum_{j=1}^{J} \left(N_j - \hat{\beta}_0(x_0) - \hat{\beta}_1(x_0)(X_j - x_0)\right)^2 K\left(\frac{|X_j - x_0|}{h}\right) \\ \cdot & \ \{1(x_0 \geq c)1(X_j > c) + 1(x_0 < c)1(X_j < c)\} \end{aligned}$$

The expression in curly brackets ensures that no observations from one side of the threshold are used to calculate density estimates on the other side.

The triangular kernel is given by the expression:

$$K(x_0) = \begin{cases} 1 - |x_0| & if\ |x_0| \leq 1 \\ 0 & otherwise \end{cases}$$

The coefficients for the local linear regression are then calculated as:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} S_0 & S_1 \\ S_1 & S_2 \end{pmatrix}^{-1} \begin{pmatrix} T_0 \\ T_1 \end{pmatrix}$$

With S_k and T_k defined as:

$$\begin{aligned} S_k &= \sum_{j=1}^{J} K\left(\frac{(X_j - x_0)}{h}\right)(X_j - x_0)^k \\ T_k &= \sum_{j=1}^{J} K\left(\frac{(X_j - x_0)}{h}\right)(X_j - x_0)^k N_j \end{aligned}$$

Consequently, the estimator at point x_0 is described by:

$$\hat{y}(x_0) = \hat{\beta}_0(x_0) = T_0 \frac{S_2 - S_1(X_j - x_0)}{S_0 S_2 - (S_1)^2}$$

The outcome of interest is then:

$$\gamma \equiv \ln \lim_{x_0 \downarrow c} y(x_0) - \ln \lim_{x_0 \uparrow c} y(x_0)$$

Define $\lim_{x_0 \downarrow c} y(x_0) = y^+$and $\lim_{x_0 \uparrow c} y(x_0) = y^-$. The estimate for a jump in the density is then:

$$\begin{aligned} \hat{\gamma} &= \ln \hat{y}^+ - \ln \hat{y}^- \\ &= \ln\left(T_0^+ \frac{S_2^+ - S_1^+(X_j - c)}{S_2^+ S_0^+ - (S_1^+)^2}\right) - \ln\left(T_0^- \frac{S_2^- - S_1^-(X_j - c)}{S_2^- S_0^- - (S_1^-)^2}\right) \end{aligned}$$

With $S_k^+ = S_k$ for $X_j > c$ and $S_k^- = S_k$ for $X_j < c$ as well as $T_k^+ = T_k$ for $X_j > c$ and $T_k^- = T_k$ for $X_j < c$.

It is shown by McCrary (2008) that the estimation bias $\sqrt{nh}\,(\hat{\gamma} - \gamma)$ is approximately normally distributed and asymptotically converges to zero under the following conditions: Everywhere except at c, the density function $y(x)$ has three continuous and bounded derivatives, $h \to 0$, $Nh \to \infty$ and $\frac{b}{h} \to 0$.[20] This leads to an approximate standard error for the estimator $\hat{\gamma}$ of:

$$\hat{\sigma}_\gamma = \sqrt{\frac{1}{hN}\frac{24}{5}\left(\frac{1}{\hat{y}^+} + \frac{1}{\hat{y}^-}\right)} \tag{4.4}$$

[20] For a proof, see Appendix I of McCrary (2008).

4.6.2 Bandwidth selection

The histogram from the first step of the specification testing procedure is taken as a starting point. Then a separate ROT bandwidth is computed for both sides of the threshold. This is done by fitting a polynomial of the fourth order to the data on each side and calculating:

$$\begin{aligned} \bar{h}_l &= \kappa \left(\bar{\sigma}_l^2 \frac{|c - X_l|}{\sum \hat{\lambda}_l''(X_j)^2} \right)^{\frac{1}{5}} \\ \bar{h}_r &= \kappa \left(\bar{\sigma}_r^2 \frac{|c - X_r|}{\sum \hat{\lambda}_r''(X_j)^2} \right)^{\frac{1}{5}} \end{aligned} \tag{4.5}$$

Where $X_l = X_0$, $X_r = X_J$. The index l describes the variables for the regression to the left of the cutoff, and index r describes those for the regression to the right. Let $\bar{\sigma}_l^2$ and $\bar{\sigma}_r^2$ be the mean squared error for the regressions on both sides of the cutoff. While $\hat{\lambda}_l''$ and $\hat{\lambda}_r''$ are the estimated second derivatives of the fourth order polynomial model.

In order to calculate the standard errors for the local linear regression as per equation 4.4, the average of both ROT bandwidths is taken. This average is used for the local linear estimator on both sides of the cutoff.

Chapter 5

Candidates' Professions and the Gender Gap in Parliaments – Experimental Evidence

5.1 Introduction

In most countries, women are underrepresented in parliaments relative to their share in the population. This gender gap can be observed across nations and at all levels of government. For instance, women occupied only 20 percent of the seats in Congress and 24.8 percent in state legislatures in the US in 2018 (CAWP, 2018), 37.4 percent in the European Parliament (European Parliament, 2017), and about 31 percent and 32 percent in the German Bundestag and the British House of Commons, respectively (IPU, 2018). This is even more remarkable since each of these entities has more women than men in the population (World Bank, 2017) and we would therefore expect female candidates to have an advantage at the polls. This underrepresentation of women has consequences for policy decisions because female politicians tend to have different policy preferences and priorities than their male colleagues (Chattopadhyay and Duflo, 2004; Clots-Figueras, 2011, 2012; Thomas, 1991). For instance, parliaments with higher shares of women assign more resources to healthcare and education (Holman, 2014).

Many factors contribute to this situation. This includes women's lower level of political ambitions (Fox and Lawless, 2010) and their higher inhibition to enter competition (Niederle and Vesterlund, 2007), as well as structural disadvantages like biased coverage in the media

(Carlin and Winfrey, 2009) and shenanigans of party leaders in favor of male candidates (see Carroll, 1994; Esteve-Volart and Bagues, 2012; Fox and Lawless, 2010; Stambough and O'Regan, 2007, among others). By contrast, the existing evidence is inconclusive about the role of the voters in this context, that is, whether there is a systematic bias against female candidates in the electorate. While some studies report small negative effects of being female (e.g. Giger et al., 2014; Sanbonmatsu, 2002), others show either no impact of candidate gender (McElroy and Marsh, 2010) or even a small advantage for women (Black and Erickson, 2003).

In this paper, we examine an aspect of voter behavior which has been overlooked in the literature so far, but may explain part of the representation gap and some of the variation in the results of these earlier studies. More specifically, we test whether voters have a preference for candidates working in typically male-dominated professions. If this is true, male candidates would possess a hidden systematic advantage over female candidates in situations in which their respective profession is either a prominent feature in the campaign or directly stated on the ballot as additional information on the candidates.[1]

For the analysis, we use data from an election experiment built into an exit poll of voters in Germany in 2014. Respondents faced a list of 30 imaginary candidates of their favorite party and were asked to select the six they would prefer to represent them. Using different information treatments, we first examine whether there is a direct gender preference among voters in the absence of other information about the candidates. To this end, we exogenously assign first names to the candidates which unambiguously indicate a certain gender. In consequence, female voters strongly prefer female candidates, while male voters seem to be indifferent towards candidate gender. The results for female voters are in line with earlier findings of same-sex preference in the literature (e.g. Dolan, 1998; Sigelman and Sigelman, 1982). Due to this strong bias among the women in our sample, female candidates enjoy a sizable bonus on average in this scenario.

In the second step, we look at the results of six different ballot versions in which we add information about the candidates' profession. More specifically, each candidate appears on two ballot versions with a male-dominated, a female-dominated and a gender-neutral profession, respectively. This way, we are able to identify the impact of the different types of professions while keeping the candidates' name, gender, and position on the list constant.

[1] The former is the case in candidate-centered elections with simple majority voting (McDermott, 2005), the latter happens in many countries around the world, including some states in the US, Germany, and Switzerland.

Our findings suggest that profession information profoundly affects the selection decision. First, individuals reveal a preference for candidates working in a profession which is typically associated with the respective individual's gender, e.g., male voters for physicists, engineers, firefighters, and metalworkers, and female voters for psychologists, elementary school teachers, elderly care nurses, and medical assistants. Second, this bias towards professions dominated by one's own gender is significantly stronger for men than women, confirming sociological research about stronger gender stereotyping among men (e.g. Miller and Budd, 1999; Miller and Hayward, 2006). As a consequence, the advantage for female candidates in the situation without profession information vanishes and even reverses into an electoral bonus for male ones, once we use our findings to simulate results with more realistic shares of male and female candidates working in male and female dominated professions. Finally, this bonus for men leads to losses in electoral ranks for female candidates and therefore to lower chances of rising to the top of the list and ending up in parliaments.

With respect to the literature, these findings suggest that studies examining gender preferences among voters in settings in which profession information does not play a role, will tend to find more positive results with respect to the electoral chances of female candidates (e.g. Baltrunaite et al., 2016; Black and Erickson, 2003; Dolan, 1998; McElroy and Marsh, 2010). While others, in which profession information is available (e.g. Chakraborty, 2012; Giger et al., 2014), will find negative results.

The paper continues as follows: In section 5.2, we present and discuss the most relevant literature on the use of information cues in low-information elections and the likely impact of profession and gender stereotypes. Then, we introduce the design of our experiment in section 5.3. Section 5.4 describes the resulting sample and shows that the random allocation of respondents to the different information treatments led to very similar comparison groups. In section 5.5 we describe how we identify the impact of gender and profession. Section 5.6 reports the empirical findings for both steps of the analysis, as well as for several robustness checks. Finally, section 5.7 discusses the potential implications of the results, limitations of this study, and possible directions for follow-up research on the topic.

5.2 Voting, professions and stereotypes

An extensive literature is dedicated to the issue of electoral results for female candidates. It is divided in three main strands.

The first one analyzes the legislative consequences of female representation among politicians. Since the number of women in parliaments and among representatives tends to be lower than that of men, many authors study the counterfactual effects on policy if gender shares were balanced. For instance, Holman (2014) and Thomas (1991) provide evidence that parliaments with stronger female representation increase government spending and assign more resources to healthcare, education and family support. Chattopadhyay and Duflo (2004) find that in India village councils with higher shares of women provide more of those public goods which benefit women. Parts of the existing literature also cover the effects of female policymakers on subsequent electoral success of female candidates. The findings are heterogeneous and depend on the country studied. Baskaran and Hessami (2018), using data from Germany, show that female candidates reach higher vote shares in local council elections if the respective mayor is a women. However, Ferreira and Gyourko (2014) find no evidence for political spillovers of female mayors in the USA.

A second group of studies is concerned with the supply of female candidates, i.e., the selection processes which determine the makeup of party lists and the characteristics of female candidates. Which therefore play an important role in determining the share of women in parliaments. Niederle and Vesterlund (2007) present evidence that women are less likely to sort themselves into competitive environments like political campaigns. Results from Fox and Lawless (2010) indicate that women are also less likely to be recruited as political candidates in the US, and if they are, it is in more competitive districts with a greater risk of losing (Carroll, 1994; Stambough and O'Regan, 2007). In consequence, Anzia and Berry (2011) as well as Ferreira and Gyourko (2014) suggest that those women who do run and get elected are more capable on average, due to the harder selection process they go through. There is mixed evidence if having more female candidates on party lists affects their representation in parliament. For instance, Campa (2011) finds no significant change in female representation after the introduction of gender quotas on ballots in Spanish municipal elections. Esteve-Volart and Bagues (2012) show evidence that this may be due to parties systematically placing female candidates on less favorable positions on the ballot, which diminishes the effects of gender quotas. However, Baltrunaite et al. (2016) exploit the introduction of new election rules in Italy in a regression discontinuity design and find evidence for strong supply side effects on the election of female representatives.

An important third strand of research, and the one to which our paper contributes the most, studies the determinants of voter demand for female candidates. Early work by Sigelman and Sigelman (1982) shows only weak gender effects among US undergraduate students, especially in comparison to the strong impact of the age of the candidates. Their experimental survey also contained information on candidate professions, but as all of them were gender neutral, it is not possible to derive any results about gender stereotyping from this setup. Huddy and Terkildsen (1993) look at gender effects from a different angle. They provide experimental evidence that voters associate typically female and male candidate traits with different areas of political expertise. Candidates with male traits are considered more competent on the topics of economics and national security, while female character traits are associated with higher competence in topics which require compassion, such as social security and education. Overall, voters seem to consider male character traits to be more important for office holders than female character traits. If this extends to male-dominated professions compared to female-dominated ones, it could explain the impression stated above that studies in environments in which professions play a role tend to find less favorable results for female candidates than in circumstances without information on occupations. Similarly, the different traits attributed to male and female candidates may also explain the finding by McDermott (1997) that female candidates fare better when running for the Democrat Party compared to the Republican Party in the US. Since their perceived competences aline much more closely with the preferences of the electorate of the former party. Going one step further in the analysis and considering the voting patterns of male and female voters separately, there is strong evidence that individuals prefer to be represented by candidates of their own gender. Such behavior is shown by both Dolan (1998) and Sanbonmatsu (2002).

The article most similar to ours in terms of experimental design and empirical approach is McDermott (2005). Using experimental survey data from state-wide races in California, she demonstrates that in low-information settings, occupational characteristics of the candidates serve as a signal for qualification and educational achievement and thus influence voter decisions. The study has two limitations with respect to our topic, however. First, the focus on races for executive offices featured only one candidate per party, which means that the party affiliation of the candidates most likely dominated the other information cues. And second, the experiment only varied whether the given profession of the candidates was revealed or not, but did not exogenously assign gender and different professions to the same candidates. Taking a closer look at gender-profession interactions and their impact on the electoral gender gap is therefore not possible with this setup.

Our paper contributes to this literature in a number of ways. First, it examines experimentally whether we can explain the different findings on the gender gap in voter preferences with the presence of profession information about the candidates. To this end, we conduct the same election experiment in the absence and presence of profession information and compare the magnitude of the respective gender gaps. Second, we investigate how different types of professions impact the electoral chances of male and female candidates separately. We focus in particular on the role of occupations that are strongly dominated by either gender and may therefore transport certain connotations. And third, we are able to use our rich data containing individual-level information on the participants, their voting behavior, and the characteristics of the candidates in order to specifically look at stereotypical and atypical gender-profession combinations and how voters react to them. To the best of our knowledge, the last two points constitute the first attempts to examine this topic with an adequately powerful experiment and sample size.

With these contributions, our paper also relates to the literature on the hiring of employees, as the context of choosing the best candidate for office resembles the selection of a new employee rather closely. In both cases, the respective principal chooses an agent out of many applicants based on a rather limited set of information about the latter. Gender preferences revealed in hiring decisions could therefore easily exist in voting decisions and vice versa. For instance, Azmat and Petrongolo (2014) conduct a comprehensive review of the international literature on gender differences in hiring and conclude that there is evidence of significant discrimination against women in high-status or male-dominated jobs and against men in female-dominated ones. If politics is considered a more male-dominated domain and more masculine competences like assertiveness and risk-taking are perceived as necessary to thrive there, this could explain a part of the observed gender gap. Additionally, if voters do not care per-se about the gender of the candidate, but appreciate candidates working in high-status or male-dominated jobs in general, this points towards possible negative spill-over effects for female candidates from the labor market into the political arena in situations in which candidate profession is an important cue.

5.3 Survey design

For the analysis of these phenomena, we use data from a large election experiment conducted as part of an exit-poll at the simultaneous elections to the EU parliament and local councils in Germany in May 2014. At a total of 28 locations in 15 different communities in the population-rich states of Baden-Württemberg and Nordrhein-Westfalen, voters were approached outside the polling stations and asked to participate in an anonymous study on voter behavior.[2] In order to obtain a representative sample of voters in these places, interviewers were instructed to approach every third person leaving the buildings. All persons

[2]The original questionnaire in German is available from the authors upon request

who consented to take the survey obtained the questionnaire and could fill it out on their own with the interviewer remaining nearby in case of questions.[3]

The questionnaire was structured in four main parts:[4] The first asked individuals about the election they just participated in. That is, which party they voted for, how many candidates of that party they knew, and how satisfied they were with their choices. The second was a hypothetical election in different versions which constitutes the core of the present study and will be explained in more detail below. The third part inquired about the experience of voting in this hypothetical election, e.g. what methods they used to select candidates and what additional information they would have liked to know about them. Finally, the fourth elicited the basic personal characteristics of the respondents, that is, their gender, age group, education, family status, and current profession.

In the election experiment, survey participants were asked to allocate six votes among a list of 30 hypothetical candidates, which they should think off as candidates from their preferred party. Thus, the setup mimics the situation of voters who have to choose between mostly unknown candidates of their preferred party on a lower institutional level. For instance in an open list election as used in many countries of Northern and Central Europe or a primary election to determine a party's candidate in a local race in the US.

Respondents were randomly assigned to a total of 16 different versions of the hypothetical election. All of them featured the same candidates, as defined by their family names, at the same ballot positions, but varied in the amount of information provided to the participants. For the purposes of this paper, we use the following eight versions as presented in figure 5.1: The first one only features the family name and the initials of the first names, such that they appear in a gender-neutral way to the participant. Version 2 spells out the whole name and thus reveals the gender of the candidates. Male and female candidates appear alternating, starting with a male candidate on the first spot. Finally, versions 3 to 8 use the same names and gender as version 2, but additionally state a profession next to the name of the candidates.

The setup of our data collection provides a number of advantages over traditional sources. Compared to surveys conducted on the phone some time after the election, it directly targets the group of people most relevant for this research at a point in time as close to the real decision-making as possible. Selection issues with regards to voting participation

[3] Bishop and Fisher (1995) have shown that filling out questionnaires oneself increases the accuracy of the answers by reducing the social desirability bias.

[4] A separate part included questions for other research projects which are not important in our context.

are minimized by drawing respondents only from those individuals who actually went to the polls. This ensures that our respondents are in the right set of mind to correctly answer questions about how they voted and how they would choose in a slightly different environment. Furthermore, a paper-based questionnaire and hypothetical election ballot is much more similar to the real act of voting and making one's cross. In particular in terms of being able to visually scan the whole list of candidates. This would not be possible in the case of phone interviews, especially when the case in question is an election with multiple candidates.

In comparison to using real election outcomes, our setup allows for a much stronger identification of the key effects of interest as the exact combinations of gender and profession of the hypothetical candidates are exogenously determined, while potentially confounding factors like ballot position, name recognition, or age are all either unknown or held constant. This approach makes it possible to precisely set the composition of candidate characteristics and perfectly isolate the effect of profession information on the voter's willingness to choose female candidates from changes in the supply of candidates. Furthermore, having information about several important characteristics of the participants enables us to examine the behavior of relevant subgroups separately, most importantly male and female voters. This is typically impossible when looking at election data due to the anonymity of the voting process.

5.4 Sample descriptives

In total, 2327 voters filled out one of the eight questionnaire versions relevant for this study. We further restrict the sample to those respondents who stated their gender and allocated all of their six votes among the candidates in the hypothetical election. Thus, we end up with 1826 respondents for the empirical analysis. Table A5.2 reports the descriptive statistics for the resulting sample. It shows that 49.8 percent of the respondents were female, the average age was around 44 years[5] and 66.8 percent had a university entrance qualification (the German "Abitur"). Finally, about 54.4 percent reported to have voted for a center-left party, which includes the Social Democratic Party (SPD), the Green Party, and the socialist Left Party (Die Linke). In terms of regional distribution, 53.1 percent of participants were interviewed in a larger city (more than 100,000 inhabitants) and 53.2 percent in the state of Baden-Württemberg.

[5] As participants were asked to indicate their age in intervals (see the questionnaire in the appendix), we use the interval means for this calculation and 75 for the group of participants over 65.

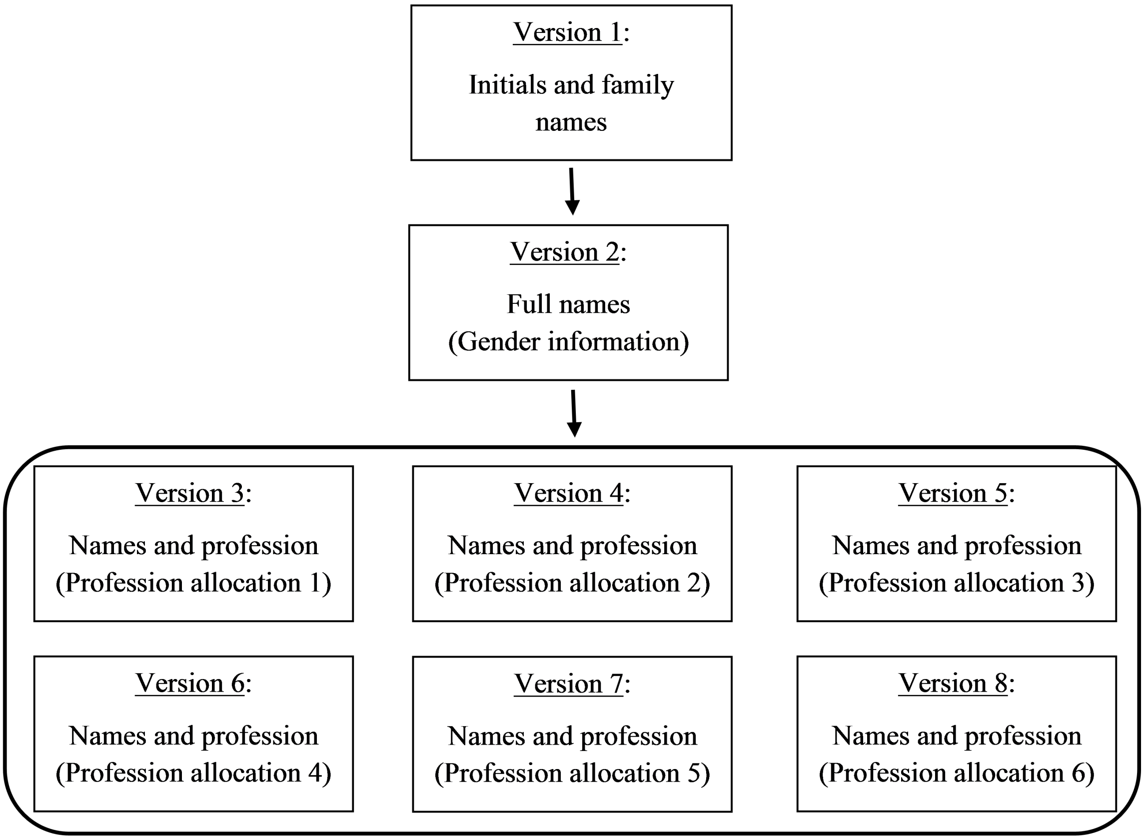


Figure 5.1: Ballot versions and informational content.

As we will often present results by voter gender in the next sections, it is also interesting to compare the participants by gender. To this end, Table A5.2 also presents the respective personal characteristics of male and female participants. The numbers show that male and female voters in the sample appear very similar. The two groups only deviate somewhat with respect to their age distribution and education level. More specifically, female voters are underrepresented in the age group between 26 and 35 years (15 vs. 20 percent among male respondents), slightly overrepresented among those aged 46 to 55 (21 vs. 18 percent), and report a somewhat smaller share of individuals with a secondary or higher degree (64 vs. 69 percent).

Finally, we want to check whether the random allocation of respondents to the different ballot versions led to very similar groups facing the various information treatments. Table A5.3 confirms that this is the case. It reports the descriptive statistics of the survey participants for each ballot version and marks those instances in bold in which the average of the respondents in one version differs significantly (on the 5 percent level) from the mean of the others. Thus, we can see that the individual groups resemble each other very much in terms of personal characteristics. Only in 8 out of 136 cells in table A5.3 do we observe significant deviations on the 5 percent level, that is, in roughly 5.6 percent of the cases. Given this wide-ranging similarity in the observable characteristics of survey participants across treatments, it seems plausible to assume that they also share the same unobservable characteristics on average and hence constitute credible counterfactuals.

5.5 Identification strategy

Given this experimental setup and the random assignment to the different ballot versions with incrementally increasing or changing informational content, we apply a three-step approach to examine the impact of profession information on the chances of female candidates to get elected. We start by focusing on the ballots without profession information to examine whether our survey participants exhibit a direct preference for either gender. Then, we look at the ballot versions displaying the candidates' occupation to check whether the inclusion of this information leads to indirect changes in the electoral prospects of female candidates. As this is going to be the case, we finally further exploit the experimental setup to test whether this effect is driven by preferences for certain gender-profession combinations. The following subsections present each of these steps in more detail.

5.5.1 Direct gender preferences

To see whether the gender gap in parliaments may be driven by direct voter preferences for male candidates, we look at the voting behavior of participants who faced hypothetical election ballots without any profession information and check whether the probability of candidate i to get one of the six votes of participant j depends on the gender of the candidate. To this end, we use ballot version 2 and estimate the following model by logit:

$$\Pr(\texttt{vote}_{ij}|X) = \Lambda(\alpha_0 + \alpha_1\texttt{female}_i + \alpha_2\texttt{rank\&name}_i) \tag{5.1}$$

Here, `female` is an indicator variable for whether the candidate appears as a woman on the ballot, i.e., appears with a traditionally female first name.[6] To take the potential influence of ballot position and family name into account, we additionally control for each candidate's average probability to receive a vote in ballot version 1 (`rank&name`). Male and female

[6]Where X is the vector of explanatory variables, in this case $X = (\texttt{female}_i, \texttt{rank\&name}_i)$.

first names are exogenously allocated to the 30 candidates on the list and the better ballot position for men is taken into account by `rank&name`. Coefficient α_1 thus represents the average electoral bonus or disadvantage female candidates possess compared to their male counterparts, i.e., the electoral gender gap.[7] In an extension to the model, we also look at possible differences in the estimated electoral gender gap by gender of the voter by including an indicator for whether voter j is female (`femvot`) and its interaction with the gender of the candidate (`female` × `femvot`).

5.5.2 Effect of profession information

In the second step, we measure whether the electoral gender gap differs in the presence of profession information on the ballot. To this end, we consider the voting behavior of participants facing ballot versions 3 to 8 In these, the same hypothetical candidates appear at exactly the same positions as in version 2, only with additional information about their respective occupation stated after their names. We pool all these observations and run the specification in equation 5.1 with this sample to see whether the estimated coefficient of `female` changes under these circumstances. If there is a large deviation from the earlier estimate obtained in the situation without profession displayed on the ballot, we can attribute this change to the inclusion of this piece of information about the candidates.

The impact of gender-profession combinations

If the presence of profession information on the ballot turns out to affect the electoral chances of male and female candidates differently on average, we want to know more about the mechanisms at work. In particular, we want to answer three interrelated questions: First, is this change in voting behavior a mere consequence of switching from one decision criterion to another (here, from gender to profession) and thus happening coincidentally, or do voters systematically connect certain professions with a specific gender and take that (un)consciously into account? Second, do certain kinds of occupations affect the electoral chances of male and female candidates differently? And third, do male and female voters react differently to gender-profession combinations?

To examine these issues, we chose the 30 professions used in our study such that they are well-known and can be characterized as one third female-dominated, gender-neutral, and male-dominated, respectively. The criterion for this sorting was the share of female workers in a

[7] We do not include any other candidate characteristics, as they are unknown to the participants by design and therefore redundant. Furthermore, all voter-specific variables do not matter here either, since we restrict the sample to participants who allocated the full six votes.

certain occupation as reported in the German Microcensus of 2010, an annually conducted representative household survey providing information on about 800,000 individuals. We define a profession as female dominated if the share of women exceeded 70 percent among its workforce in that year, as gender neutral if it lay between 40 and 60 percent, and as male dominated if it was below 30 percent. Table 5.1 shows the selected professions and their respective share of female workers. To prevent that differences in the skill level of the selected occupations influence our analysis, we also ensured that half the professions in each gender category can be considered as high and low skill employment. The corresponding values for the share of higher educated individuals (defined as having obtained the *Abitur*, the general qualification for university entrance in Germany) among the workforce in the respective profession are also displayed in table 5.1, with thresholds of above 70 and below 30 percent, respectively. Thus, we end up with six distinct groups of five professions each, distinguished by gender dominance and skill level.

We distributed the 30 occupations to the 30 candidates on each list such that every candidate features a distinct profession on any given ballot and none of them appears twice. Furthermore, we exogenously varied the allocation of professions to candidates over the six ballot versions which contain this information. More specifically, every candidate appears with an occupation from a different gender-dominance and skill-level group in each version.[8,9] Given this exogenous variation of professions across candidates and ballot versions and since participants were randomly allocated to the different ballot versions, we can use this setup to identify the true effects of working in one type of profession compared to another as well as how this varies with the gender of both the candidate and the voter, respectively.

Type of occupation

We start with a very simple model, in which we only focus on the type of occupation. That is, we regress the probability of candidate i to get one of the six votes of participant j on a constant and indicators for male- and female-dominated occupations, controlling for nothing except ballot position and family name:

$$\begin{aligned}\Pr\left(\mathtt{vote}_{ij}|X\right) = \Lambda\Big(&\beta_0 + \beta_1\mathtt{female\text{-}dominated}_i + \beta_2\mathtt{male\text{-}dominated}_i \\ &+ \beta_3\mathtt{rank\&name}_i\Big)\end{aligned} \tag{5.2}$$

[8]Table A5.1 in the appendix displays which profession was shown for each candidate across ballot versions.

[9]Two candidates (numbers 7 and 9 from the list) slightly deviated from this rule due to an apparent accident in the allocation. In consequence, one of them (number 9) appears with the same profession in two ballot versions.

Table 5.1: Fraction of female workers and graduates from Gymnasium in the selected professions.

		Profession	Fraction of ... Female	Gym. grad.
Female-dominated	Higher educated	Psychologist	0.726	0.972
		Elem. schoolteacher	0.781	0.966
		Pharmacist	0.734	0.964
		Social pedagogue	0.718	0.884
		Bookseller	0.726	0.738
	Lower educated	Elderly care nurse	0.864	0.264
		Medical assistant	0.990	0.208
		Textile cleaner	0.853	0.160
		Cleaner	0.890	0.147
		Hairdresser	0.905	0.131
Gender-neutral	Higher educated	Teacher	0.569	0.998
		Lawyer	0.569	0.993
		Dentist	0.424	0.989
		Physician	0.471	0.987
		Local public servant	0.492	0.835
	Lower educated	Inkeeper	0.418	0.327
		Postal worker	0.507	0.234
		Retailer	0.565	0.215
		Cook	0.575	0.234
		Confectioner	0.557	0.154
Male-dominated	Higher educated	Physicist	0.218	0.987
		Construction engineer	0.167	0.971
		Electrical engineer	0.044	0.956
		Software developer	0.135	0.837
		Computer scientist	0.137	0.767
	Lower educated	Firefighter	0.003	0.300
		Carpenter	0.019	0.203
		Farmer	0.222	0.179
		Metal worker	0.145	0.137
		Painter	0.063	0.088

Source: German Microcensus 2010, own calculations.

This way, we obtain the average effects of appearing with a female- or male-dominated profession on the ballot on the candidates' electoral chances relative to the situation in which they would be identified as working in a gender-neutral profession. If coefficients β_1 and β_2 turn out to be statistically insignificant, we can conclude that the presence of profession information changes the electoral chances of male and female candidates either only coincidentally or through some other channel than gender perceptions. If they are statistically significant, on the other hand, we can interpret them as evidence that the voters systematically connect certain professions to a gender and use this information for their decision.

Profession preference by gender of voters

Next, we examine whether male and female voters react differently to the inclusion of candidate professions, i.e., we look at differences in the preferences for individual profession types between male and female voters. Each time, we introduce an indicator for whether the voter/candidate is female and interact it with the indicators for the profession types. Given the established fact in the literature that voters are more likely to vote for someone similar to them (e.g. Cutler, 2002; Sigelman and Sigelman, 1982), we would expect male voters to favor male-dominated professions and female voters to prefer female-dominated professions on average. With respect to the gender of the candidate, the effect may go both ways. If voters follow traditional views of what men and women are supposed to do, this would trigger better electoral prospects for male candidates working in male professions and female candidates in female professions. However, if atypical gender-profession combinations catch more attention, inspire sympathy or respect, the results could go into the other direction.

Typical and atypical gender-profession combinations

Finally, we combine the two perspectives and consider the impact of different types of occupations by candidate and voter gender. To this end, we rerun the analysis from the previous section, but separately for male and female voters. This will enable us to see whether male or female voters differ with respect to their support for traditional or atypical gender-profession combinations. As previous studies show that men tend to adhere more strongly to traditional views about the appropriate professions for men and women (Miller and Hayward, 2006), it is possible that this behavior plays a role in elections as well.

We also estimate an additional specification which includes a variable which indicates if the candidate and voter have a similar profession, and a variable which indicates if they have similar profession and the same gender. Evidence from theoretical and experimental psychology strongly suggests that voters prefer candidates which display socioeconomic characteristics similar to their own.[10] We expect that this preference for similarity also applies to professions. The inclusion of the indicators allows us to analyze if such similarity effects are present in the data and estimate their strength. In order to separate the effects of profession and gender, we control for average vote probabilities per rank from ballot version 2.[11]

5.6 Empirical analysis

5.6.1 Direct gender effects

Direct gender effects without profession information

In a first step, we quantify the effects of candidate gender on vote share which are caused by the shift from ballot version 1 to ballot version 2. These are direct gender effects as differences in vote share between otherwise identical candidates of varying gender and reflect differences in voter preferences between male and female candidates. Gender is visible to voters in the form of unambiguous first names. Version 1 only contains information on candidate surnames, while version 2 adds information cues about first names.

The first column of table 5.3 shows the vote share effect of being a female candidate, reported in odds ratios.[12] Table A5.4 reports the same effects in percentage points. All standard errors are clustered at the voter level in order to account for correlation between candidate choices by the same voter.

Female candidates show a significant vote share bonus relative to male candidates. Con-

[10]For instance, Piliavin (1987) shows experimental evidence for similarity effects in age, race and sex. Goldstein and Gigerenzer (2002) explore the heuristic mechanism which leads to similarity preferences.

[11]The same results can be obtained by controlling for rank averages from ballot 1, as in equation (5.1), and additionally controlling for candidate gender.

[12]Abbreviating $\Pr(y = 1)$ as $\Pr(y)$, the odds ratio is defined as

$$OR = \frac{\Pr(y|x = 1)/(1 - \Pr(y|x = 1))}{\Pr(y|x = 0)/(1 - \Pr(y|x = 0))}$$

and shows how many times more likely an outcome of $y = 1$ is relative to $y = 0$, if x is equal to 1. If the estimated odds ratio for the female candidate indicator is larger than 1, female candidates are more likely to receive the vote than male candidates. If it is smaller, female candidates are less likely to receive the vote. The odds ratio shows the odds of receiving the vote for a female candidate as a share of the male candidate's odds. For small values of $\Pr(y = 1)$, the odds ratio is a good approximation of the relative probability of an outcome. In this case, an odds ratio for the female indicator of, for example, 1.05 can be interpreted as an approximately 5 percent higher chance of female candidates receiving the vote.

ditional on ballot position and influence of their family name, they receive approximately 33 percent more votes and their likelihood of receiving the vote is 4.47 percentage points higher than that of males. This is a substantial vote share bonus when considering that purely random voting would lead to a probability of 20 percent to receive each vote. Male candidates have an average probability of receiving the vote of 19.2 percent while female candidates, on average, receive 21.1 percent. Being female therefore is no drawback for candidates in list elections. On the contrary, voters even seem to have a baseline preferance for female candidates.

As an additional measure of electoral success, we calculate average list rank gains or losses for candidate groups when switching from ballots with less information to ballots with more information.[13] These measures of rank gain reflect the effect on electoral achievement which was caused by the introduction of the next information cue. When considering list rank gains as a measure of electoral success, female candidates climb an average of 1.47 ranks relative to their position in the list of ballot 1.

In the second columns of tables 5.3 and A5.4, the effect of being a female candidate is interacted with voter gender. The base category, against which the vote share of all other voter and candidate combinations are measured, are male candidates and male voters. Relative to this base category the vote share of female candidates and male voters is not significantly different, as shown in the third row of table 5.3. However in the second row of table 5.3 female candidates are approximately 33 percent more likely to receive the vote of a female voter than male candidates are likely to receive the vote of a man. The fourth row of table 5.3 shows that male candidates are approximately 23 percent less likely to receive the vote of a female voter than that of a male voter. The cross table of odds ratios between all voter and candidate categories (table 5.4) reinforces the notion that the positive vote share difference for female candidates is driven by female voters. For instance, female candidates are 27.5 percent more likely to receive the vote of a women than that of a man. While female voters display significant preferences for candidate gender, male voters are indifferent with respect to candidate gender. Thus, direct gender effects are driven by female voter

[13] In each ballot version, the number of votes determines the final rank each candidate achieves in the election outcome. For instance, moving from ballot version 1 to ballot 2 we can calculate for each candidate the rank gain by subtracting the final rank in ballot version 2 from the final rank which the candidate achieves in ballot version 1. Taking the average of rank gains and losses over the subgroups of male and female candidates, we obtain the effect of gender on list rank. Average rank gains for candidates of certain profession types (interacted with gender) are calculated analogous in the move from ballot version 2 to ballot versions 3 to 8. List rank gains provide a very intuitive measure of electoral success. However, list rank gains of individual candidates also depend on the election results of other candidates. Because of the discontinuous nature of list ranks the results are less precise than measured vote shares.

bias against male candidates. The probability of receiving the vote of a man does not differ significantly between male and female candidates, which indicates that men do not use gender as a cognitive shortcut for candidate skill. In this respect, our findings align with those of Dolan (1998) who shows that female voters are more sensitive to candidate gender. However, indifference towards gender does not imply that men vote randomly if gender is the only information cue available. When asked about the method by which they allocate their votes in this low information setting, the majority of respondents stated that they based their decisions on names and the position in the candidate list (compare table 5.2). These statements are confirmed by the highly significant coefficient on the list rank effect in table 5.3, which indicates that male voters base their voting decision in this setting on list rank and surname.

Table 5.2: Methods used to allocate the votes, by information treatment and gender

	Information stated on the ballot (Treatment)					
	None		Gender		Gender & Profession	
Fraction stating this particular type of method among those who used one (Open question, multiple answers possible)						
Voter gender	Males	Females	Males	Females	Males	Females
Names	**0.448**	**0.569**	**0.472**	**0.458**	0.040	0.028
Foreigners	0.207	0.120	0.139	0.125	0.018	0.012
First 6 on the list	0.207	0.120	0.056	0.104	0.033	0.007
Randomly	0.000	0.060	0.056	0.041	0.016	0.002
Profession	0.067	0.020	0.000	0.000	**0.687**	**0.719**
Age	0.100	0.000	0.056	0.146	0.007	0.012
Gender	0.034	0.020	0.361	0.408	0.093	0.155
Family status	0.034	0.000	0.000	0.000	0.000	0.007
Education	0.000	0.000	0.000	0.021	0.111	0.102
Representativeness	0.000	0.000	0.083	0.042	0.084	0.106
Similar to oneself	0.000	0.000	0.000	0.021	0.029	0.039

Notes: (1) This table shows the fraction of voters indicating to have used the respective method among those who declared they had allocated their votes in a particular way. In order to solicit unbiased answers, this question was asked in open format, i.e., without predetermined alternatives. (2) The method used by the most individuals is marked in bold for each ballot version.

Direct gender effects with profession

Next, we test if the direct effect of candidate gender on vote share depends on the availability of profession information on the ballot. To this end, we estimate specification 1 on the pooled sample of ballot versions 3 to 8. Results are shown in columns three of table 5.3 and table A5.4. When candidate professions are included in the information on the ballot there is no vote share effect for female candidates. The odds ratio for the female candidate indicator of 0.998 is close to one and statistically insignificant. Gender based candidate choice by voters, which leads to a substantial vote share advantage for female candidates in ballot 2, appears to be replaced by a separate set of voter preferences which is heavily influenced by professions. In the questionnaire about the method which they chose to allocate votes, 70 percent of voters with ballots 3 to 8 state that candidate profession was the most important factor that they based their decision on (table 5.2). These new preferences do not favor female candidates.

Now we interact the female candidate indicator with voter gender, replicating regression specification 2 on a sample from ballots 3 to 8. The aim is to evaluate if male and female voters respond differently to the inclusion of profession information on the ballot. Results are displayed in columns four of table 5.3 and table A5.4. When professions are unknown, male voters are indifferent towards candidate gender while female voters prefer female candidates. With known professions we see a different pattern. Male voters no longer balance their votes between male and female candidates, instead they are about 15.3 percent less likely to give the vote to a female candidate. Female voters, on the other hand, are 15.6 percent less likely to give the vote to a male candidate. Profession information by itself can't be the cause of these shifts in preferences because profession information cues are perfectly balanced between candidates of both genders by experimental design. Therefore it is the interaction of profession information and candidate gender which leads to gender bias in voting by both male and female voters. In the following section we explore the influence of profession information cues and their interactions with candidate gender in detail.

Table 5.3: Direct gender effects

	Ballot 2	Ballot 2	Ballots 3-8	Ballots 3-8
Female cand.	1.335^{***} (0.0881)		0.998 (0.0257)	
F.cand.*F.vot.		1.335^{***} (0.0801)		1.002 (0.0295)
F.cand.*M.vot.		1.047 (0.0875)		0.0847^{***} (0.0295)
M. cand.*F.vot.		0.768^{***} (0.0517)		0.844^{***} (0.0295)
Rank effects ballot v. 1	Yes	Yes	No	No
Rank effects ballot v. 2	No	No	Yes	Yes
Base Male cand.	0.192		0.199	
Base Mcand, Mvot		0.211		0.212

Note: Exponentiated coefficients; Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Effects in odds ratios show how much more likely a candidate is to receive the vote, relative to the base category. Under fully random voting, the base probability would be 0.2 for each candidate.

Table 5.4: Cross table of direct gender effects

Base	F. cand.*M. vot.	M. cand.*F. vot.	M. cand.*M. vot.
F. cand.*F. vot.	1.275***	1.738***	1.335***
F. cand.*M. vot.		1.363***	1.047
M. cand.*F. vot.			0.768***

Note: Effects in odds ratios, $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5.6.2 Indirect gender effects

Indirect gender effects through profession type

Table 5.5: Raw profession effects

Profession type	Male dominated	Neutral	Female dominated
Probability of receiving the vote	21.6%	18.6%	19.8%

Note: Under fully random voting, each occupation category would receive 20% of all votes.

In this chapter we analyze indirect gender effects which are transmitted through profession information cues. Men and women are not evenly distributed between professions. Some professions are heavily dominated by males, e.g. engineering, manual labor and construction. Other professions are almost entirely dominated by women. These professions are often found in education and healthcare. Many professions are gender neutral in the sense that equal numbers of men and women pursue them. For instance lawyers and retail workers. Ballot versions 3 to 8 add information cues on professions of each of the three types defined by gender dominance. Each ballot contains equal numbers of male and female candidates in each profession so that vote share effects of gender dominated professions are independent of candidate gender. Within each profession type, professions are balanced by required qualification. This ensures that gender dominance and not perceived education levels determine voter preferences for each profession type. Even when not controlling for list ranks, the profession types differ in their probability of receiving the vote. In terms of raw probabilities, male dominated professions are more likely to receive the vote than female dominated ones, and both are more likely to be voted for than gender neutral professions (table 5.5).

This pattern continues in the results from a logit model which controls for list rank effects. We show the effects of each profession type in figure 5.2. As in the case of raw probabilites, we see evidence of strong profession type effects. Male dominated professions enjoy a significant vote share bonus and are 20 percent more likely to receive the vote than neutral professions. Female dominated professions are significantly more likely to be voted for than neutral professions (8 percent), but are significantly less likely to be voted for than male professions.[14] In terms of list ranks, candidates with male professions gain, on average, 1.85

[14] The difference between the vote share bonuses for candidates in male dominated and female dominated professions is highly significant with a t-test p-value of 0.0023.

rank positions. Candidates with female professions gain 0.17 ranks and neutral professions lose 1.68 list ranks (table 5.8).

This clear ordering of categories establishes that gender dominance plays an important role in voter valuation of professions. If male dominated professions act as information shortcuts for typically male character traits like assertiveness and toughness, as suggested by Huddy and Terkildsen (1993) and Lemkau (1983, 1984), it would explain voter preference for such professions. However in that case we would expect both male voters and female voters to prefer candidates in male professions. As we show in figure 5.3 and table A5.5, only male voters prefer male professions while female voters are indifferent between male dominated and neutral professions. Therefore the vote share bonus for male professions is unlikely to be driven by inferred political ability of such professions.

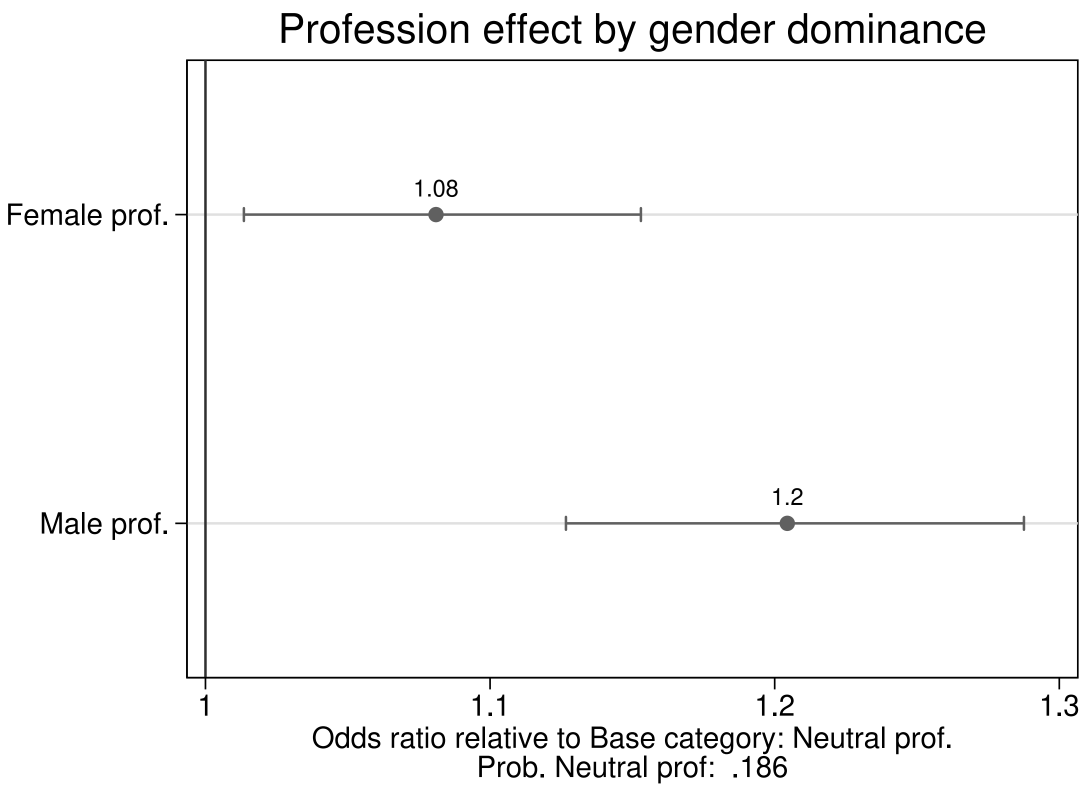


Figure 5.2

Profession type effects by voter gender

We now interact gender dominated professions with voter gender using data from ballot versions 3 to 8. Therefore, we quantify how voters value professions which are typically held by members of their own gender.

Figure 5.3 , table A5.5 and table 5.6 show that female voters clearly prefer female dominated professions to male or neutral professions. They are almost 30 percent more likely to vote for candidates in female dominated professions than for candidates in male or neutral professions. Conversely, male voters bestow a large vote share bonus (almost 40 percent) upon candidates in male dominated professions, relative to neutral professions. They are also significantly less likely to vote for candidates in female professions compared to those in male or neutral professions. It is therefore clear that voters favor candidates in professions which are dominated by their own gender. This gender biased profession effect is more pronounced among male voters, which stands in contrast with the lack of direct gender preferences which male voters display when professions are unknown (compare section 5.6.1). Male voters do not let gender directly influence their voting decision but show strong indirect gender preferences through the occupation channel. They favor candidates with male dominated professions and show a dislike for candidates in female dominated professions. Female voters on average prefer female professions but their biases are weaker than those of male voters. The vote share bonus they give to candidates in female dominated professions is lower than the bonus given by male voters to male professions. Female voters are also indifferent between male dominated and neutral professions.

Table 5.6: Cross table of profession effects by voter gender

Base	F. vot.*Male prof.	F. vot.*Neutral prof.	M. vot.*Male prof.	M. vot.*Female prof.	M. vot.*Neutral prof.
F. vot.*Female prof.	1.284***	1.317***	0.962***	1.452***	1.279***
F. vot.*Male prof.		1.026	0.72***	1.133***	1.000
F. vot.*Neutral prof.			0.703***	1.104**	0.970
M. vot.*Male prof.				1.569***	1.380***
M. vot.*Neutral prof.					0.879**

Note: Effects in odds ratios

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

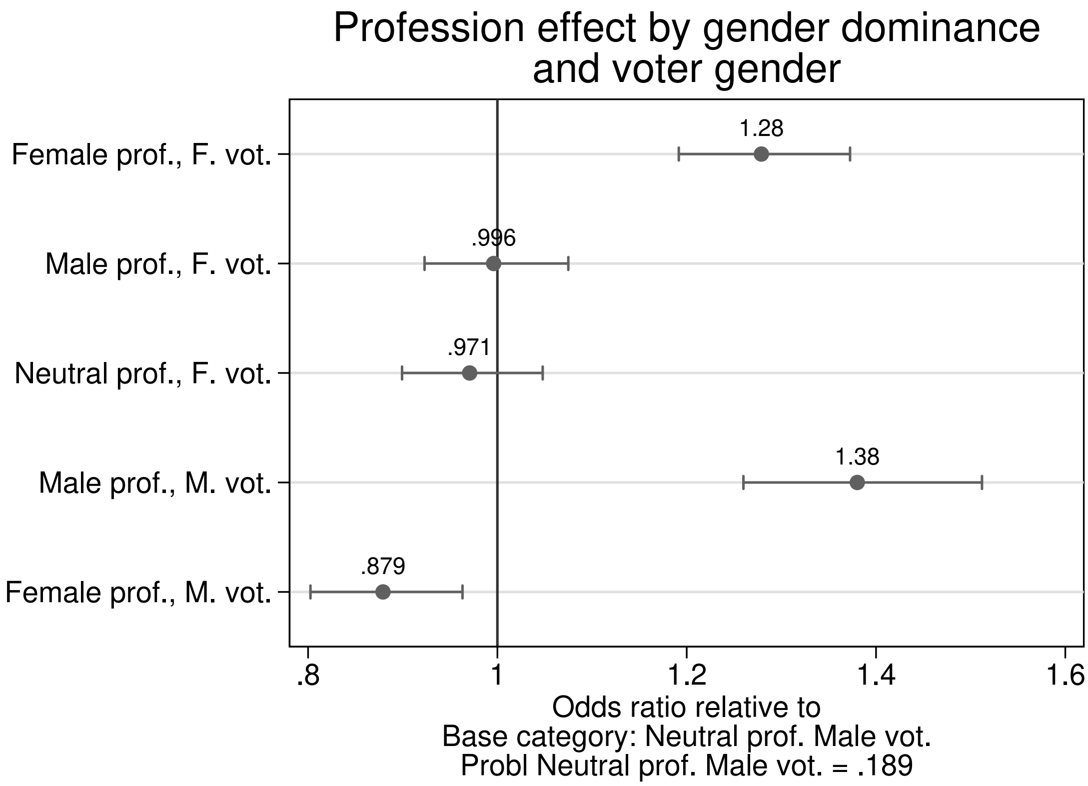


Figure 5.3

Profession type effects by candidate gender

In the next step, we interact gender dominated profession types with candidate gender. The interaction of candidate genders and profession describes whether or not the candidate conforms to the stereotypes of "men doing men's work" or "women doing women's work". By testing if voters value the profession types differently for male and female candidates we can quantify the effects of stereotypical gender-profession combinations on vote share. Since voters use information cues as cognitive shortcuts, candidates with stereotypical gender and profession combinations might enjoy a vote share bonus if voters perceive such stereotypical candidates as more predictable or more similar to themselves. The latter has been shown to be an important determinant of voter choice, since voters gravitate towards candidates which are close to them in terms of sociodemographic distance (Cutler, 2002). Results are displayed in figure 5.4 table A5.6 and cross-table 5.7. Candidates who adhere to the stereotype of men in male dominated professions are preferred over male candidates in other professions. They are approximately 32.5 percent more likely to be voted for than male candidates in neutral professions, and 27 percent more likely than male candidates in female dominated professions.

Female candidates enjoy a significant vote share advantage if they adhere to the stereotype of women in women's professions compared to women in neutral jobs. However, female candidates in male professions also gain a vote share advantage over those in neutral professions. The vote share advantages of female candidates in both male and female dominated professions are of similar size. There is no statistically significant difference in the probability of receiving the vote between females in female or male dominated professions (compare column one, row one of table 5.7). While female candidates in female professions enjoy a vote share advantage of 17 percent over females in neutral professions, females in male professions have an advantage of 13 percent over those in neutral professions. For candidates in neutral professions, their gender does not influence the vote share in any significant way. These effects are partially reflected in the way that candidates who adhere to stereotypes gain list ranks. Table 5.8 shows that stereotypical male candidates gain an average of 3.33 list ranks when moving from ballot version 2 to ballot versions 3 to 8. Atypical male candidates on the other hand lose 0.4 ranks. Stereotypically female candidates gain 0.067 ranks and female candidates in male professions gain 0.367 ranks. While stereotypically female candidates receive more votes, their positions on ballots 2 and 3 to 8 are such that despite their increased vote share they don't always manage to surpass the threshold required for higher ranks.

Table 5.7: Cross table of profession effects by candidate gender

Base	F. cand.*Male prof.	F. cand.*Neutral prof.	M. cand.*Male prof.	M. cand. Female prof.	M. cand. Neutral prof.
F. cand.*Female prof.	1.030	1.139***	0.889***	1.137***	1.163***
F. cand.*Male prof.		1.107**	0.864***	1.105**	1.130***
F. cand.*Neutral prof.			0.780***	0.998	1.020
M. cand.*Male prof.				1.279***	1.308***
M. cand.*Female prof.					1.022

Note: Effects in odds ratios

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

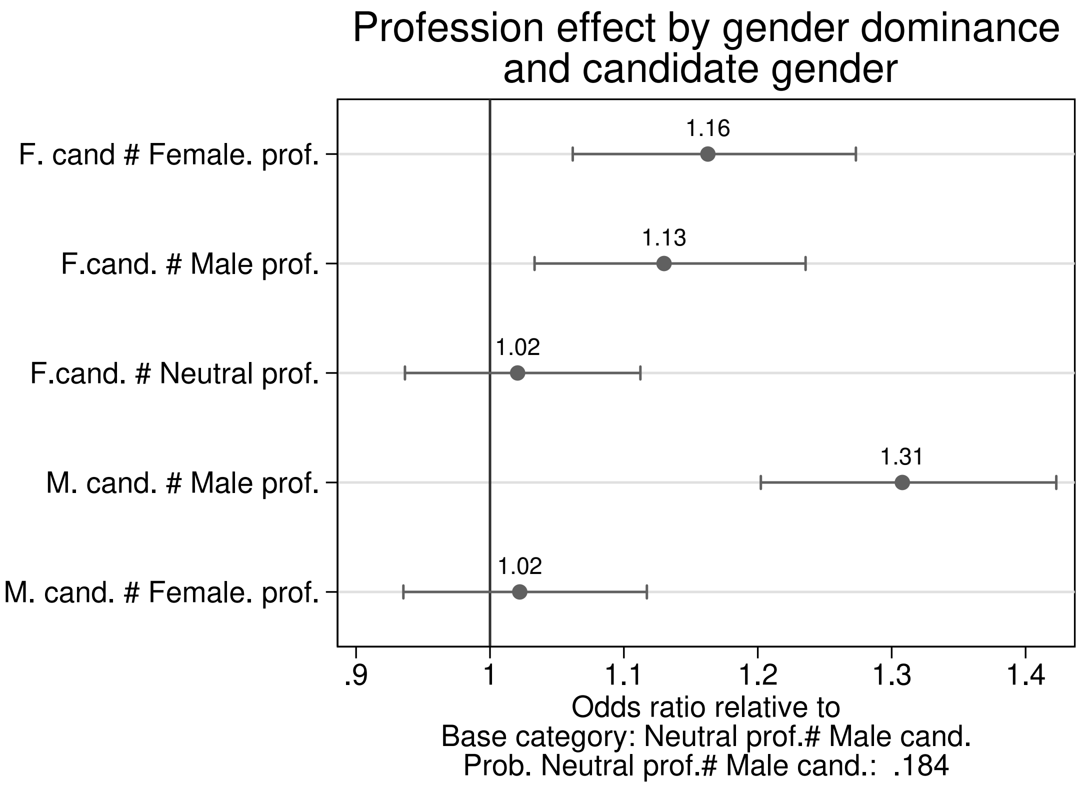


Figure 5.4

Profession type effects by candidate and voter gender

Next, we interact profession types with candidate gender and report results separately for male and female voters. Results are displayed in figure 5.5 and table A5.7. Male voters have a clear preference for candidates who adhere to the stereotype of men in male dominated professions. These candidates are approximately 41 percent more likely to receive the male voter's vote than male candidates in neutral professions. Male candidates in female professions, or female candidates in female and neutral professions get even fewer votes from male voters.

Female voters strongly prefer stereotypical female candidates. They award a vote share bonus of roughly 72 percent to female candidates in female professions, relative to the base category of male candidates in neutral professions. However, this base category receives the lowest vote share from female voters, which slightly overemphasitzes the bonus for stereotypical female candidates. All other candidate categories have significantly higher probabilities of being voted for than the base category, with odds ratios ranging between

Table 5.8: List rank gains

	A. List ranks gained for candidates in each profession type					
Profession type	Male dominated		Neutral		Female dominated	
	0.167		-1.683		1.850	
	B. List ranks gained for each profession type, stratified by candidate gender					
Profession type	Male dominated		Neutral		Female dominated	
Candidate gender	Male	Female	Male	Female	Male	Female
	3.333	0.367	-1.300	2.067	-0.400	0.067

Notes: Panel A shows the gain and loss in list ranks when profession information is made available to the voters, averaged over all candidates of the respective profession type. Panel B shows the same list rank gains as in panel A, but averaged over subgroups defined by profession type and candidate gender.

1.17 and 1.32. Still, female candidates in female professions receive by far the most votes from female voters.

We conclude that gender stereotypes are advantageous for the vote share of male candidates, but that their role in determining the vote share of female candidates is more complex. Females gain a vote share advantage from having typically female professions, just like male candidates gain from typically male professions, although the advantage for women is somewhat smaller. In both cases, this advantage is driven by voters of the respective gender. But female candidates also profit from male dominated professions, almost to the same degree as from female professions. However, while the vote share bonus for stereotypical female candidates is driven by female voters, the bonus for non-stereotypical females is driven by male voters who show a baseline preference for candidates in male professions.

Vote share bonuses for stereotypical candidates might appear through two channels. First, the information cues of candidate gender and gender typical profession complement each other to create the image of a stereotypical person. Stereotyping allows the voters to infer additional characteristics about the candidate, which he or she might not know but which are part of the stereotype (compare Rahn (1993)). This in turn makes such candidates more predictable which might positively influence voter choice in low information contexts which always carry a lot of uncertainty for the voter. Second, voters are more likely to vote for candidates which are similar to themselves, as shown by Cutler (2002). Additional evidence from theoretical and experimental psychology strongly suggests that voters prefer candidates which display socioeconomic characteristics similar to their own.[15] We find that similarity between voters and candidates with respect to gender, profession and the combination thereof leads to strong positive vote share effects. In table 5.9 we show that voters have a very high chance of giving the vote to candidates who have the same profession as themselves. The odds of giving the vote to a candidate with the same profession are several times higher than with other candidates.[16] Even more so if they share the same gender. By definition, voters of both genders more often pursue occupations dominated by their own gender. Therefore female voters are more likely to have the same profession and gender as stereotypical female candidates and the same holds for male voters, which in turn leads to higher vote shares for stereotypical candidates. Note, however, that the vote shares of profession/candidate combinations do not vary by much when controlling for profession/gender similarity of candidate and voter (table 5.9). Similarity between voter and candidate is therefore a significant predictor of vote probability, but only explains a small part of the vote share differences between stereotypical and non-stereotypical candidates. This is mainly because the number of voters who have exactly the same profession as a candidate is fairly small in our sample. Only 4 percent of voters share a profession with a candidate and only 2 percent share both profession and gender.

It is likely that voters prefer candidates who are similar to themselves in a wider sense, which would vastly expand the number of voters who encounter similar candidates on ballots. For instance, they might prefer candidates who work in fields which are related to their profession. A male voter who works as a carpenter might, for example, prefer candidates who are craftsmen in other fields. The expanded similarity channel might therefore still contribute substantially to the stereotype preferences which we observe.

Extension: Simulated vote shares and rank gains under realistic distribution of professions

In the previous section we have shown that voter preference for male candidates in male dominated professions is stronger than for female candidates in female dominated professions. Conversely, voters give less votes to male candidates in female dominated professions

[15] Piliavin (1987) shows experimental evidence for similarity effects in age, race and sex. Goldstein and Gigerenzer (2002) explore the heuristic mechanism which leads to similarity preferences.

[16] Such high odds ratios can no longer be reliably interpreted as percentage differences but instead serve to illustrate the relative strength of the effect. We show the results in odds ratios for comparison of profession stereotype effects with figure 5.4.

Table 5.9: Candidate to voter similarity

Same profession indicator	5.089***	(1.140)
Same gender indicator	1.181***	(0.0306)
Same profession and gender	1.582	(0.505)
F. cand.*Female prof.	1.174***	(0.0536)
F.cand.*Male prof.	1.148***	(0.0527)
F.cand.*Neutral prof.	1.028	(0.0450)
M. cand.*Male prof.	1.317***	(0.0568)
M. cand.*Female. prof.	1.031	(0.0467)
Rank effects ballot v. 2	Yes	
P_val_same_jointly	0	

Note: Exponentiated coefficients; Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Effects in odds ratios show how much more likely a candidate
is to receive the vote,relative to the base category.
Under fully random voting, the base probability would be 0.2
for each candidate.

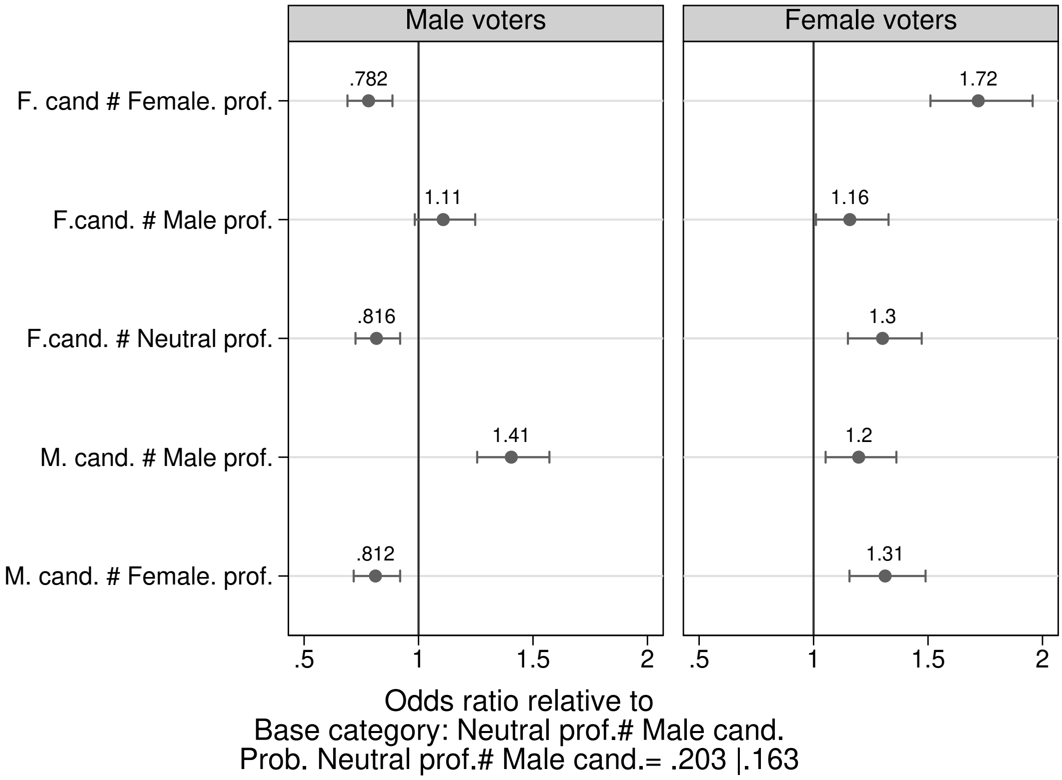


Figure 5.5

than to female candidates in male dominated professions. Aggregating over all professions we find no vote share advantage for either gender when profession information cues are available to the voter. Estimating the gender specific vote share effect for ballot versions 3 to 8, which contain the profession information cues, shows that female candidates receive a statistically insignificant 0.02 percent higher vote share. Their probability of receiving the vote is only 0.037 percentage points higher. Compared to the ballot versions without profession information cues, in which female candidates receive 33 percent more votes than male candidates, this is a substantial drop in vote share. While voters on average like to vote for female candidates, this bonus vanishes when candidate professions are known.[17]

The ballots in the experimental setting contain equal numbers of candidates for each combination of gender and profession-type. However, by definition, vastly more men pursue male dominated professions and more women work in female dominated ones. Such a gender

[17]As shown in section 5.6.1, female voters vastly prefer female candidates while male voters are indifferent with regards to candidate gender.

imbalance within professions will be reflected in the makeup of party lists. Therefore the candidates in real elections do not display the same balanced distribution of professions across genders as the ballots in our experiment. The more pronounced vote share bonus of stereotypically male candidates (31 percent) relative to other candidate types will then lead to a reduced vote share for the aggregate of female candidates. Stereotypically female candidates only gain a bonus of 16 percent, non-stereotypical females 13 percent and non-stereotypical males receive no vote share bonus at all. A reduction in the aggregate female vote share is simply a mechanical consequence of increasing the numbers of stereotypical candidates on the ballot relative to the number of non-stereotypical ones.[18] The number of male candidates who earn the highest vote share increases while the number of male candidates who receive no bonus decreases. For females however, the number of candidates who earn a vote share bonus of 16 percent increases, and the number of those who gain a 13 percent bonus decreases. Since those two vote shares are similar, they largely balance each other out and the aggregate female vote share does not change. In order to quantify the effect of realistic profession distributions on aggregate female vote share, we simulate vote results for different distributions stereotypical candidates by reweighting the share of candidate gender/profession combinations. We show the results for the observed experiment and for three different simulated distributions. Without reweighting the share of female candidates in each profession type (male dominated, female dominated, neutral) is 50 percent. The first simulated distribution has a share of 70 percent female candidates within female dominated professions and a share of 30 percent females within male dominated professions, with the reverse holding for males. The second distribution has a female share of 80 percent in female dominated professions and 20 percent in male dominated ones and the third has shares of 90 percent and 10 percent, respectively. These shares all represent realistic distributions of gender typical professions. In reality, women make up between 99 and $\approx$ 70 percent of workers in female dominated professions and between 0.03 and 22 percent in the male dominated ones (compare table 5.1). Vote share results for the three simulations are given in table 5.10. They show that under realistic distributions, female candidates suffer a vote share disadvantage of -3.2, -4.7 and -6.1 percent, respectively, for the three simulated distributions. Compared to the vote share bonus which female candidates receive if professions are unknown, this is a drastic decline. Driving these changes in vote share are voter preferences for stereotypical professions, not a preference for candidate gender in

[18]Here we assume that no general equilibrium effects affect the vote share. That is, voter preferences for candidate types are independent of candidate supply as long as at least one candidate of each type is present on the ballot.

itself. Voter predilection for professions overrides their preferences for female candidates, causing the swing in voter demand. When asked directly about the criteria they chose to select candidates, the vast majority of voters declared that they based their voting decision on profession information if that information was known (table 5.2). When professions were unknown, many voters selected candidates for their name and gender. More specifically, male voters who previously were indifferent to gender, rather give their vote to male candidates in male dominated professions than to any other candidates. This leads to a reduction in the female vote share bonus from 33 percent to almost zero in our sample and to a female vote share penalty of up to 6.1 percent when professions are realistically distributed by gender.

Table 5.10: Simulated profession distributions

		Gender information	Gender and profession information			
		Benchmark	Experiment	Simulation 1	Simulation 2	Simulation 3
Share of female candidates in...	female professions	n.a.	0.5	0.7	0.8	0.9
	neutral professions	n.a.	0.5	0.5	0.5	0.5
	male professions	n.a.	0.5	0.3	0.2	0.1
Share of male candidates in...	female professions	n.a.	0.5	0.3	0.2	0.1
	neutral professions	n.a.	0.5	0.5	0.5	0.5
	male professions	n.a.	0.5	0.7	0.8	0.9
Share of stereotypical female candidates on the list		n.a.	0.167	0.233	0.267	0.300
Share of stereotypical male candidates on the list		n.a.	0.167	0.233	0.267	0.300
Advantage of stereotypical male candidates over stereotypical females		n.a.	9.72 %	9.72 %	9.72 %	9.72 %
Aggregate vote share bonus of female candidates relative to male candidates		33 %	0.02 %	-3.20 %	-4.70 %	-6.10 %
Average rank gain of female candidates relative to previous ballot		1.467	-0.544	-0.584	-0.604	-0.624

5.7 Conclusion

In this study, we examine whether information about the profession of candidates running for public office affects the electoral gender gap, i.e., the difference in probability to get elected between men and women. To this end, we use data from an election experiment built into an exit-poll of voters at the elections to the EU parliament in Germany in 2014. We exploit the random allocation of participants into various information treatments to obtain unbiased estimates for direct and indirect gender effects. Our findings suggest that voters, especially women, are inclined to lend their support to female candidates if there are no other cues about the candidates' characteristics and positions. Once the candidates' occupations are revealed, however, the voters use these as their main selection criterion. This leads to shrinking support for female candidates and may even turn into an electoral bonus for male candidates. This remarkable change is driven by male voters who strongly favor stereotypical male candidates working in male-dominated professions. The magnitude of this effect may even be underestimated, as we abstained from including professions indicating manager or supervisor positions, which are typically male-dominated, and houseman or housewife, in which women are overrepresented.

In essence, these results reveal an interesting pattern in the voting behavior of male voters, boiling down to a position of: "I don't care about the gender of the candidates, as long as they are competent, and I believe people working in male professions are just that." In a sense, focusing on profession leads to same-sex preference. This has implications for gender equity policy. It means that the still existing segmentation of parts of the labor market spills-over to the political arena as well. In consequence, any progress of women on the labor market will also lead to progress in addressing the electoral gender gap. That is, policies to promote gender quality at work may also increase female representation in parliaments and councils in the medium and long-run. It also means that there is no baseline-preference for male candidates. Which implies that policy measures which aim to improve female representation in parliaments do not necessarily stand in opposition with voter preferences.

The findings in this paper are not transferable to every political context, however. They are derived and relevant for elections with a low information environment in which the gains from inferred information from stereotyping are much greater than in situations in which the candidates are well known. Typical examples for such low information elections are open lists and primary elections with many candidates belonging to the same party, or any election in which the candidates are relatively unknown, typically on lower institutional levels. By contrast, high-profile races for important political offices in which only a handful of prominent candidates compete against each other are unlikely to be influenced by gender stereotyping. Besides, candidates at the national level tend to be full-time career politicians who only marginally differ in terms of profession anyway. Further research on the strength of gender-profession stereotypes in different contexts may therefore focus on determining which share of voters use heuristics for their voting decision across different election systems, institutional levels, and numbers of candidates running.

5.8 Appendix

5.8.1 Tables from section 5.4

Table A5.1: Candidates profession across ballot versions.

Pos.	Family name	First name	Profession 1	Profession 2	Profession 3	Profession 4	Profession 5	Profession 6
1	Gillen	Arnold	Physicist	Physician	Elem. schoolteacher	Firefighter	Innkeeper	Textile cleaner
2	Heyer	Regina	Elderly care nurse	Electrical engineer	Lawyer	Pharmacist	Carpenter	Innkeeper
3	Armrein	Karl	Confectioner	Textile cleaner	Physicist	Physician	Social pedagogue	Carpenter
4	Tesch	Iris	Social pedagogue	Farmer	Retailer	Elderly care nurse	Phyiscist	Phyiscian
5	Höhne	Otto	Metalworker	Cook	Medical assistant	Construction engineer	Dentist	Psychologist
6	Lötz	Margarete	Laywer	Pharmacist	Carpenter	Innkeeper	Hairdresser	Software developer
7	Peters	Bernd	Hairdresser	Bookseller	Farmer	Retailer	Elderly care nurse	Computer scientist
8	Gussmann	Ute	Construction engineer	Dentist	Psychologist	Metalworker	Cook	Medical assistant
9	Kilic	Mehmet	Teacher	Software developer	Teacher	Bookseller	Firefighter	Retailer
10	Kunde	Hildegard	Firefighter	Innkeeper	Hairdresser	Computer scientist	Physician	Elem. schoolteacher
11	Berger	Martin	Psychologist	Firefighter	Confectioner	Medical assistant	Electrical engineer	Local public servant
12	Silbernagel	Marianne	Postal worker	Cleaner	Electrical engineer	Local public servant	Elem. schoolteacher	Firefighter
13	Gorges	Hans-Peter	Innkeeper	Hairdresser	Software developer	Dentist	Psychologist	Farmer
14	Kleine	Erika	Bookseller	Painter	Postal worker	Cleaner	Computer scientist	Lawyer
15	Bernsen	Karl-Heinz	Painter	Postal worker	Cleaner	Electrical engineer	Lawyer	Pharmacist
16	Block	Sille	Dentist	Psychologist	Firefighter	Confectioner	Textile cleaner	Electrical engineer
17	Weber	Daniel	Electrical engineer	Lawyer	Pharmacist	Painter	Postal worker	Cleaner
18	Schenzer	Bärbel	Cleaner	Construction engineer	Dentist	Psychologist	Farmer	Confectioner
19	Lütticken	Reinhardt	Physician	Elem. schoolteacher	Painter	Postal worker	Cleaner	Construction engineer
20	Propach	Inge	Farmer	Confectioner	Elderly care nurse	Software developer	Teacher	Bookseller
21	Altenburg	Jürgen	Elem. schoolteacher	Carpenter	Innkeeper	Hairdresser	Construction engineer	Dentist
22	Greiner	Waltraud	Retailer	Medical assistant	Construction engineer	Teacher	Bookseller	Painter
23	Leksen	Walter	Medical assistant	Computer scientist	Local public servant	Social pedagogue	Metalworker	Cook
24	Benz	Barbara	Computer scientist	Local public servant	Social pedagogue	Carpenter	Retailer	Hairdresser
25	Schüttle	Heinrich	Cook	Elderly care nurse	Computer scientist	Lawyer	Pharmacist	Metalworker
26	Rudnick	Julia	Pharmacist	Metalworker	Cook	Textile cleaner	Software developer	Teacher
27	Nawak	Thomas	Carpenter	Retailer	Textile cleaner	Physicist	Local public servant	Social pedagogue
28	Block	Christiane	Local public servant	Social pedagogue	Metalworker	Cook	Medical assistant	Physicist
29	Usleber	Johannes	Software developer	Teacher	Bookseller	Farmer	Confectioner	Elderly care nurse
30	Lochner	Susanne	Textile cleaner	Physicist	Physician	Elem. schoolteacher	Painter	Postal worker

Table A5.2: Descriptive statistics for the sample of voters

	Full sample		Male voters		Female voters		Male-Female difference
	mean	sd	mean	sd	mean	sd	p-values
Female voter	0.50	0.50					
Age below 26	0.21	0.41	0.20	0.40	0.22	0.41	0.292
Age 26-35	0.18	0.38	0.20	0.40	0.15	0.36	0.005
Age 36-45	0.15	0.36	0.15	0.36	0.15	0.35	1.000
Age 46-55	0.19	0.40	0.18	0.38	0.21	0.41	0.106
Age 56-65	0.14	0.35	0.14	0.34	0.14	0.35	1.000
Age over 65	0.13	0.34	0.14	0.34	0.13	0.34	0.530
Secondary degree	0.67	0.47	0.69	0.46	0.64	0.48	0.023
City	0.53	0.50	0.54	0.50	0.52	0.50	0.393
Baden Wuerttemberg	0.53	0.50	0.52	0.50	0.54	0.50	0.393
Voted for left-leaning party	0.54	0.50	0.53	0.50	0.56	0.50	0.200
Observations	1826		916		910		

Table A5.3: Randomization across ballot versions

Version	1	2	3	4	5	6	7	8
Sample size	213	222	239	255	247	253	246	238
Female	**0.612**	0.479	0.524	0.442	0.466	0.479	0.511	0.489
	(0.49)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Age 16-25	**0.134**	0.212	0.224	**0.260**	0.204	0.199	0.229	0.187
	(0.34)	(0.41)	(0.42)	(0.44)	(0.40)	(0.40)	(0.42)	(0.39)
Age 26-35	0.189	0.157	0.155	0.186	0.196	0.220	0.157	0.143
	(0.39)	(0.36)	(0.36)	(0.39)	(0.40)	(0.42)	(0.36)	(0.35)
Age 36-45	0.144	0.180	0.164	0.124	0.132	0.144	0.178	0.126
	(0.35)	(0.38)	(0.37)	(0.33)	(0.34)	(0.35)	(0.38)	(0.33)
Age 46-55	0.209	0.161	0.203	0.190	0.200	0.191	0.195	0.209
	(0.41)	(0.37)	(0.40)	(0.39)	(0.40)	(0.39)	(0.40)	(0.41)
Age 56-65	0.179	**0.207**	0.125	0.103	0.132	**0.093**	0.131	0.143
	(0.38)	(0.41)	(0.33)	(0.30)	(0.34)	(0.29)	(0.34)	(0.35)
Age 66+	0.144	**0.083**	0.129	0.136	0.136	0.153	0.110	**0.191**
	(0.35)	(0.28)	(0.34)	(0.34)	(0.34)	(0.36)	(0.31)	(0.39)
Currently married	0.547	0.519	0.560	0.521	0.536	0.485	0.511	0.555
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Children [y/n]	0.555	0.525	0.534	0.471	0.545	0.468	0.515	**0.591**
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)
No. of children	1.144	0.977	1.034	0.975	1.081	0.940	1.030	1.117
	(1.26)	(1.15)	(1.17)	(1.25)	(1.25)	(1.33)	(1.32)	(1.20)
Education low	0.385	0.307	0.328	0.295	0.340	0.358	0.297	0.316
	(0.49)	(0.46)	(0.47)	(0.46)	(0.47)	(0.48)	(0.46)	(0.47)
A-level	0.265	0.326	0.310	0.361	0.302	0.259	0.360	0.294
	(0.44)	(0.47)	(0.46)	(0.48)	(0.46)	(0.44)	(0.48)	(0.46)
University degree	0.350	0.367	0.362	0.344	0.357	0.384	0.343	0.390
	(0.48)	(0.48)	(0.48)	(0.48)	(0.48)	(0.49)	(0.48)	(0.49)
Vote share SPD	0.275	0.293	0.287	0.243	0.292	0.266	0.278	0.302
	(0.43)	(0.45)	(0.44)	(0.42)	(0.45)	(0.43)	(0.44)	(0.45)
Vote share CDU	0.253	0.236	0.263	0.292	0.257	0.253	0.316	0.285
	(0.42)	(0.42)	(0.43)	(0.45)	(0.43)	(0.43)	(0.46)	(0.44)
Large city	0.521	0.532	0.515	0.510	0.534	0.561	0.528	0.521
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
State of BW	0.526	0.527	0.556	0.514	0.502	0.534	0.520	0.538
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)

Note: Mean values of the respective variables by ballot version. Clustered standard errors in parentheses. Means which deviate significantly (5 % level) from random allocation in bold. Only 8 out of 136 cells show significant deviation.

5.8.2 Tables from section 5.6.1

Table A5.4: Direct gender effects

	Ballot 2	Ballot 2	Ballots 3-8	Ballots 3-8
Female cand.	0.0447***		-0.000371	
	(0.0101)		(0.00412)	
Fcand*Fvot		0.0466***		0.000262
		(0.00998)		(0.00387)
Fcand*Mvot		0.00720		-0.0258***
		(0.0131)		(0.00528)
Mcand*Fvot		-0.0392***		-0.0264***
		(0.00956)		(0.00391)
Rank effects ballot v. 1	Yes	Yes	No	No
Rank effects ballot v. 2	No	No	Yes	Yes
Base Male cand.	0.192		0.199	
Base Mcand*Mvot		0.211		0.212

Notes: Standard errors in parentheses

The coefficients show percentage point changes in the probability of receiving the vote.

Under fully random voting, the base probability would be 0.2 for each candidate.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5.8.3 Tables from section 5.6.2

Table A5.5: Profession effect by gender dominance and voter gender

Female prof.*F. vot.	0.0410***	(0.00628)
Male prof.*F. vot.	-0.000624	(0.00617)
Neutral prof.*F. vot.	-0.00469	(0.00613)
Male prof.*M. vot.	0.0544***	(0.00830)
Female prof.*M. vot.	-0.0200***	(0.00702)
Rank effects ballot v. 2	0.218***	(0.0251)
Base Neutral prof. M. vot.	0.189	

Notes: Standard errors in parentheses
The coefficients show percentage point changes in the probability of receiving the vote. Under fully random voting, the base probability would be 0.2 for each candidate.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5.6: Profession effect by gender dominance and candidate gender

F. cand.*Female. prof.	0.0248***	(0.00782)
F.cand.*Male prof.	0.0200***	(0.00762)
F.cand.*Neutral prof.	0.00327	(0.00706)
M. cand.*Male prof.	0.0451***	(0.00757)
M. cand.*Female. prof.	0.00352	(0.00729)
Rank effect ballot v. 2	0.219***	(0.0251)
Base M. cand.*Neutral prof.	0.184	

Notes: Standard errors in parentheses
The coefficients show percentage point changes in the probability
of receiving the vote. Under fully random voting, the base probability
would be 0.2 for each candidate.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5.7: Profession effect by gender dominance and candidate gender

	Male voters	Female voters
F. cand.*Female prof.	-0.0372***	0.0953***
	(0.00915)	(0.0126)
F.cand.*Male prof.	0.0165*	0.0240**
	(0.00999)	(0.0117)
F.cand.*Neutral prof.	-0.0309***	0.0440***
	(0.00886)	(0.0110)
M. cand.*Male prof.	0.0576***	0.0296***
	(0.0103)	(0.0112)
M. cand.*Female prof.	-0.0317***	0.0455***
	(0.00924)	(0.0113)
Rank effects ballot v. 2	0.213***	0.223***
	(0.0358)	(0.0348)
Base M. cand.*Neutral prof.	0.203	0.163

Notes: Standard errors in parentheses
The coefficients show percentage point changes in the probability
of receiving the vote. Under fully random voting, the base probability
would be 0.2 for each candidate.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5.8: Profession type effect - fixed effect model

	Profession type	by voter gender	by candidate gender
Female prof.	0.0118**		
	(0.00476)		
Male prof.	0.0297***		
	(0.00474)		
Female prof.*M. vot.		-0.0191***	
		(0.00660)	
Female prof.*F. vot.		0.0402***	
		(0.00670)	
Neutral prof.*F. vot.		-0.00450	
		(0.00669)	
Male prof.*M. vot.		0.0541***	
		(0.00659)	
Male prof.*F. vot.		-0.000710	
		(0.00669)	
M. cand.*Female. prof.			0.00263
			(0.00684)
F. cand.*Female. prof.			-0.0729***
			(0.0171)
F.cand.*Neutral prof.			-0.0937***
			(0.0170)
M. cand.*Male prof.			0.0434***
			(0.00691)
F.cand.*Male prof.			-0.0777***
			(0.0172)
Constant	0.186***	0.188***	0.265***
	(0.00336)	(0.00466)	(0.0128)

Notes: Standard errors in parentheses

The coefficients show percentage point changes in the probability of receiving the vote.

Under fully random voting, the base probability would be 0.2 for each candidate.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6 References

Acemoglu, D. and Autor, D. (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4b of *Handbooks in Economics*, chapter 12, pages 1044–1166. North Holland, Amsterdam.

Albrecht, J., van Vuuren, A., and Vroman, S. (2009). Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the Netherlands. *Labour Economics*, 16:383–396.

Alvarez, M. and Hall, T. E. (2006). Controlling democracy: The principal-agent problems in electoral administration. *The Policy Studies Journal*, 34:491–510.

Andrews, D. W. and Schafgans, M. M. (1998). Semiparametric estimation of the intercept of a sample selection model. *The Review of Economic Studies*, 65(3):497–517.

Antonczyk, D., DeLeire, T., and Fitzenberger, B. (2018). Polarization and rising wage inequality: Comparing the US and Germany. *Econometrics*, 6(2):20.

Antonczyk, D., Fitzenberger, B., and Leuschner, U. (2009). Can a task-based approach explain the recent changes in the German wage structure? *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik)*, 229(2-3):214–238.

Antonczyk, D., Fitzenberger, B., and Sommerfeld, K. (2010). Rising wage inequality, the decline of collective bargaining, and the gender wage gap. *Labour Economics*, 17:835–847.

Anzia, S. F. and Berry, C. R. (2011). The Jackie (and Jill) Robinson effect: Why do congresswomen outperform congressmen? *American Journal of Political Science*, 55(3):478–493.

Arellano, M. and Bonhomme, S. (2017). Quantile selection models with an application to understanding changes in wage inequality. *Econometrica*, 85(1):1–28.

Arulampalam, W. (2001). Is unemployment really scarring? Effects of unemployment experiences on wages. *The Economic Journal*, 111(475):585–606.

Autor, D. (2013). The “task approach” to labor markets: An overview. *Journal for Labour Market Research*, 46(3):185–199.

Autor, D., Levy, F., and Murnane, R. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118(4):1279–1333.

Autor, D. H., Katz, L. F., and Kearney, M. S. (2008). Trends in US wage inequality: Revising the revisionists. *The Review of Economics and Statistics*, 90(2):300–323.

Azmat, G. and Petrongolo, B. (2014). Gender and the labor market: What have we learned from field and lab experiments? *Labour Economics*, 30:32–40.

Baltrunaite, A., Casarico, A., Profeta, P., and Savio, G. (2016). Let the voters choose women. *CESifo Working Paper Series*, 5693.

Baskaran, T. and Hessami, Z. (2018). Does the election of a female leader clear the way for more women in politics? *American Economic Journal: Economic Policy*, 10(3):95–121.

Bauer, T., Bonin, H., Goette, L., and Sunde, U. (2007). Real and nominal wage rigidities and the rate of inflation: Evidence from west german micro data. *The Economic Journal*, 117(524):F508–F529.

BBC News (2012). No early end to Mexico’s election stand-off. `http://www.bbc.com/news/world-latin-america-18700953`. Accessed: 2014-11-13.

Beblo, M., Wolf, E., et al. (2002). How much does a year off cost? Estimating the wage effects of employment breaks and part-time periods. *Brussels Economic Review*, 45(2):191–217.

Biewen, M., Fitzenberger, B., and Lazzer, J. d. (2018). The role of employment interruptions and part-time work for the rise in wage inequality. *IZA Journal of Labor Economics*, 7(1):10.

Biewen, M., Fitzenberger, B., and Seckler, M. (2019). Counterfactual quantile decompositions with selection correction taking into account huber/melly (2015): An application to the german gender wage gap. *Unpublished Manuscript, University of Tbingen and Humboldt-University Berlin.*

Biewen, M. and Seckler, M. (2017). Changes in the German wage structure: Unions, internationalization, tasks, firms, and worker characteristics. *IZA Discussion Paper*, No. 10763.

Bishop, G. and Fisher, B. S. (1995). "Secret ballots" and self-reports in an exit-poll experiment. *Public Opinion Quarterly*, 59(4):568–588.

Black, J. H. and Erickson, L. (2003). Women candidates and voter bias: Do women politicians need to be better? *Electoral Studies*, 22(1):81–100.

Blundell, R., Costa Dias, M., Meghir, C., and Shaw, J. (2016). Female labor supply, human capital, and welfare reform. *Econometrica*, 84(5):1705–1753.

Bollinger, C., Ziliak, J. P., and Troske, K. R. (2011). Down from the mountain: Skill upgrading and wages in Appalachia. *Journal of Labor Economics*, 29(4):819–857.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *The American Economic Review*, 77(4):531–553.

Borowczyk-Martins, D. and Lalé, E. (2016). Employment adjustment and part-time jobs: The US and the UK in the great recession. *IZA Discussion Paper*, No. 9847.

Brenke, K. and Müller, K. U. (2013). Gesetzlicher Mindestlohn – Kein verteilungspolitisches Allheilmittel. *DIW Wochenbericht*, 39:3–17.

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the USA: A quantile regression approach. *Journal of Applied Econometrics*, 13:1–30.

Buchinsky, M. (2001). Quantile regression with sample selection: Estimating women's return to education in the U.S. *Empirical Economics*, 26:87–113.

Burda, M. C. and Hunt, J. (2011). What explains the German labor market miracle in the great recession? *Brookings Papers on Economic Activity*, pages 273–336.

Burda, M. C. and Mertens, A. (2001). Estimating wage losses of displaced workers in Germany. *Labour Economics*, 8(1):15–41.

Caliendo, M., Schröder, C., and Wittbrodt, L. (2019). The causal effects of the minimum wage introduction in germany–an overview. *German Economic Review*.

Campa, P. (2011). Gender quotas, female politicians and public expenditures: Quasi-experimental evidence. *ECONPUBBLICA Working Paper*, No. 157.

Campbell, T. (2005). *Deliver the Vote: A History of Election Fraud, an American Political Tradition - 1742-2004*. Carrol & Graf Publishers, New York.

Card, D., Heining, J., and Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality. *The Quarterly Journal of Economics*, 128(3):967–1015.

Carlin, D. B. and Winfrey, K. L. (2009). Have you come a long way, baby? Hillary Clinton, Sarah Palin, and sexism in 2008 campaign coverage. *Communication Studies*, 60(4):326–343.

Caro, R. A. (1990). *The Years of Lyndon Johnson: Means of Ascent*. Alfred A. Knopf, Inc., New York.

Carroll, S. J. (1994). *Women as candidates in American politics*. Indiana University Press, Indianapolis.

Caughey, D. and Sekhon, Jasjeet, S. (2011). Election and the regression discontinuity design: Lessons from close U.S. house races, 1942-2008. *Political Analysis*, 19:385–408.

CAWP (2018). Center for American Women and Politics: Women in the U.S. Congress 2018. `http://www.cawp.rutgers.edu/women-us-congress-2018`. Accessed: 2018-11-13.

Chakraborty, L. (2012). Electoral success in Swedish municipal councils: The role of occupation and politicians' characteristics. Uppsala University.

Chattopadhyay, R. and Duflo, E. (2004). Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica*, 72(5):1409–1443.

Chen, S. and Khan, S. (2003). Semiparametric estimation of a heteroskedastic sample selection model. *Economic Theory*, 19:1040–1064.

Cheng, M.-Y. (1997). A bandwidth selector for local linear density estimators. *The Annals of Statistics*, 25(3):1001–1013.

Cheng, M.-Y., Fan, J., and Marron, J. S. (1997). On automatic boundary corrections. *The Annals of Statistics*, 25(4):1691–1708.

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268.

Clots-Figueras, I. (2011). Women in politics: Evidence from the Indian states. *Journal of public Economics*, 95(7):664–690.

Clots-Figueras, I. (2012). Are female leaders good for education? Evidence from India. *American Economic Journal: Applied Economics*, 4(1):212–44.

Connolly, S. and Gregory, M. (2009). The part-time pay penalty: Earnings trajectories of British women. *Oxford Economic Papers*, 61:76–97.

Cook, T. D. (2008). "Waiting for life to arrive": A history of the regression-discontinity design in psychology, statistics and economics. *Journal of Econometrics*, 142:636–654.

Cutler, F. (2002). The simplest shortcut of all: Sociodemographic characteristics and electoral choice. *Journal of Politics*, 64(2):466–490.

Das, M., Newey, W. K., and Vella, F. (2003). Nonparametric estimation of sample selection models. *The Review of Economic Studies*, 70(1):33–58.

Dell, M. (2015). Trafficking networks and the Mexican drug war. *American Economic Review*, 105(6):1738–1779.

D'Haultfoeuille, X., Maurel, A., and Zhang, Y. (2014). Extremal quantile regressions for selection models and the black-white wage gap. *IZA Discussion Papers*, No. 8256.

Dickert-Conlin, S. and Chandra, A. (1999). Taxes and the timing of births. *Journal of Political Economy*, 107(1):161–177.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages. *Econometrica*, 64(5):1001–1044.

Dinardo, J. and Lee, D. S. (2004). Economic impacts of unionization on private sector employers: 1984-2001. *The Quarterly Journal of Economics*, 119:1383–1441.

Dolan, K. (1998). Voting for women in the "year of the woman". *American Journal of Political Science*, pages 272–293.

Dustmann, C., Fitzenberger, B., Schönberg, U., and Spitz-Oener, A. (2014). From sick man of Europe to economic superstar: Germany's resurgent economy. *Journal of Economic Perspectives*, 28(1):167–88.

Dustmann, C., Ludsteck, J., and Schönberg, U. (2009). Revisiting the German wage structure. *The Quarterly Journal of Economics*, 124:843–881.

Dustmann, C. and Schönberg, U. (2011). Expansions in maternity leave coverage and childrens long-term outcomes. *American Economic Journal: Applied Economics*, 4:190–224.

Edin, P.-A. and Gustavsson, M. (2008). Time out of work and skill depreciation. *ILR Review*, 61(2):163–180.

Edler, S., Jacobebbinghaus, P., and Liebig, S. (2015). Effects of unemployment on wages: differences between types of reemployment and types of occupation. *SFB 882 Working Paper Series*, 51. Bielefeld.

Esteve-Volart, B. and Bagues, M. (2012). Are women pawns in the political game? Evidence from elections to the spanish senate. *Journal of Public Economics*, 96(3-4):387–399.

European Parliament (2017). Men and Women distribution. `http://www.europarl.europa.eu/elections2014-results/en/gender-balance.html`. Accessed: 2018-11-13.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, volume 66 of *Monographs on Statistics and Applied Probability.* Chapman and Hall, London.

Felbermayr, G., Baumgarten, D., and Lehwald, S. (2014a). Increasing wage inequality in Germany: What role does global trade play? *Global Economic Dynamics, Bertelsmann Stiftung.* Gütersloh.

Felbermayr, G., Hauptmann, A., and Schmerer, H.-J. (2014b). International trade and collective bargaining outcomes: Evidence from German employer–employee data. *The Scandinavian Journal of Economics*, 116(3):820–837.

Ferreira, F. and Gyourko, J. (2014). Does gender matter for political leadership? The case of us mayors. *Journal of Public Economics*, 112:24–39.

Fitzenberger, B. (1999). *Wages and Employment Across Skill Groups: An Analysis for West Germany.* Physica/Springer, Heidelberg.

Fitzenberger, B. (2012). Expertise zur Entwicklung der Lohnungleichheit in Deutschland. Arbeitspapier, Sachverständigenrat zur Begutachtung der Gesamtwirtschaftlichen Entwicklung.

Fitzenberger, B., Osikominu, A., and Völter, R. (2006). Imputation rules to improve the education variable in the IAB employment subsample. *Schmollers Jahrbuch*, 126(3):405–436.

Fitzenberger, B., Steffes, S., and Strittmatter, A. (2016). Return-to-job during and after parental leave. *The International Journal of Human Resource Management*, 27(8):803–831.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4A of *Handbooks in Economics*, chapter 1, pages 2–97. North Holland, Amsterdam.

Fox, R. L. and Lawless, J. L. (2010). If only they'd ask: Gender, recruitment, and political ambition. *The Journal of Politics*, 72(2):310–326.

Gans, J. S. and Leigh, A. (2008). Born on the first of july: An (un)natural experiment in birth timing. *Journal of Public Economics*, 93:246–263.

Gartner, H. (2005). The imputation of wages above the contribution limit with the German IAB employment sample. *FDZ Methodenreport*, No. 2.

Gernandt, J. and Pfeiffer, F. (2007). Rising wage inequality in Germany. *Jahrbücher für Nationalökonomie und Statistik*, 227(4):358–380.

Giger, N., Holli, A. M., Lefkofridi, Z., and Wass, H. (2014). The gender gap in same-gender voting: The role of context. *Electoral Studies*, 35:303–314.

Goldstein, D. G. and Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109(1):75.

Gregory, M. and Jukes, R. (2001). Unemployment and subsequent earnings: Estimating scarring among British men 1984–94. *The Economic Journal*, 111(475):607–625.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression discontinuity design. *Econometrica*, 69:201–209.

Hauser, C. and Holusha, J. (2006). Problems lead 8 states to extend some voting hours. *The New York Times - International Herald Tribune*. `http://www.nytimes.com/2006/11/08/world/americas/08iht-web.1108day.3441243.html?_r=1&emc=eta1`.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Heckman, J. (1990a). Varieties of selection bias. *The American Economic Review*, 80(2):313–318.

Heckman, J. J. (1981). Statistical models for discrete panel data. *Structural analysis of discrete data with econometric applications*, pages 114–178.

Heckman, J. J. (1990b). *Econometrics*, chapter Selection Bias and Self-selection, pages 201–224. Palgrave Macmillan UK, London.

Heckman, J. J. and Smith, J. A. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives*, 9(2):85–110.

Heckman, J. J. and Vytlacil, E. J. (2007a). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. *Handbook of econometrics*, 6:4779–4874.

Heckman, J. J. and Vytlacil, E. J. (2007b). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of econometrics*, 6:4875–5143.

Holman, M. R. (2014). Sex and the city: Female leaders and spending on social welfare programs in US municipalities. *Journal of Urban Affairs*, 36(4):701–715.

Huber, M. and Melly, B. (2015). A test of the conditional independence assumption in sample selection models. *Journal of Applied Econometrics*, 30(7):1144–1168.

Huddy, L. and Terkildsen, N. (1993). Gender stereotypes and the perception of male and female candidates. *American Journal of Political Science*, pages 119–147.

Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3):933–959.

Imbens, G. W. and Lemieux, T. (2008). Regression Discontinuity Designs: A guide to practice. *Journal of Econometrics*, 142:615–635.

IPU (2018). Inter-parliamentary Union. `http://archive.ipu.org/wmn-e/classif.htm`. Accessed: 2018-11-13.

Jenkins, J. A. (2004). Partisanship and contested election cases in the house of representatives, 1789-2002. *Studies in American Political Development*, 18:112–135.

Kimball, D. C., Kropf, M., and Battles, L. (2006). Helping America vote? Election administration, partisanship, and provisional voting in the 2004 election. *Election Law Journal*, 5:447–461.

Klein, R. and Spady, R. (1993). An efficient semiparametric estimator for binary response models. *Econometrica*, 61(2):387–421.

Kohn, K. (2006). Rising wage dispersion, after all! The German wage structure at the turn of the century. *IZA Discussion Paper*, No. 2098.

Lalive, R. (2008). How do extended benefits affect unemployment duration? A regression discontinity approach. *Journal of Econometrics*, 142:785–806.

Lee, D. S. (2001). The electoral advantage to incumbency and voters' valuation of politicians' experience: A regression discontinuity analysis of elections to the U.S. house. *National Bureau of Economic Research Working Paper*, 8441.

Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. house elections. *Journal of Econometrics*, 142:675–697.

Lee, D. S. and Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 142:655–674.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48:281–355.

Lehoucq, F. (2003). Electoral fraud: Causes, types, and consequences. *Annual Review of Political Science*, 6:233–256.

Lemieux, T. (2006). Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *The American Economic Review*, 96(3):461–498.

Lemkau, J. P. (1983). Women in male-dominated professions: Distinguishing personality and background characteristics. *Psychology of Women Quarterly*, 8(2):144–165.

Lemkau, J. P. (1984). Men in female-dominated professions: Distinguishing personality and background features. *Journal of Vocational Behavior*, 24(1):110–122.

Ljungqvist, L. and Sargent, T. (1998). The European unemployment dilemma. *Journal of Political Economy*, 106(3).

Lott Jr., J. R. (2001). Documenting unusual declines in republican voting rates in Florida's western panhandle counties in 2000. *Public Choice*, 123(3). Unpublished, American Enterprise Institute, Crime Prevention Research Center.

Ludwig, J. and Miller, D. L. (2005). Does head start improve children's life chances? Evidence from a regression discontinuity design. *National Bureau of Economic Research Working Paper*, 11702.

Machado, C. (2017). Unobserved selection heterogeneity and the gender wage gap. *Journal of Applied Econometrics*, 32(7):1348–1366.

Machado, J. A. F. and Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20(4):445–465.

Manning, A. and Petrongolo, B. (2008). The part-time pay penalty for women in Britain. *The Economic Journal*, 118(526):28–51.

McCann, J. A. (1998). Mexicans react to electoral fraud and political corruption: An assessment of public opinion and voting behavior. *Electoral Studies*, 17(4):483–503.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design, a density test. *Journal of Econometrics*, 142(2):698–714.

McDermott, M. L. (1997). Voting cues in low-information elections: Candidate gender as a social information variable in contemporary United States elections. *American Journal of Political Science*, 41(1):270–283.

McDermott, M. L. (2005). Candidate occupations and voter information shortcuts. *The Journal of Politics*, 67(1):201–219.

McElroy, G. and Marsh, M. (2010). Candidate gender and voter choice: Analysis from a multimember preferential voting system. *Political Research Quarterly*, 63(4):822–833.

Melly, B. (2005). Public-private sector wage differentials in Germany: Evidence from quantile regression. *Empirical Economics*, 30:505–520.

Melly, B. (2006). Public and private sector wage distributions controlling for endogenous sector choice. Discussion Paper, University of St. Gallen.

Miller, L. and Budd, J. (1999). The development of occupational sex-role stereotypes, occupational preferences and academic subject preferences in children at ages 8, 12 and 16. *Educational Psychology*, 19(1):17–35.

Miller, L. and Hayward, R. (2006). New jobs, old occupational stereotypes: Gender and jobs in the new economy. *Journal of Education and Work*, 19(1):67–93.

Möller, J. (2016). Lohnungleichheit - Gibt es eine Trendwende? *Wirtschaftsdienst*, 96(1):38–44.

Moretti, E. (2013). Real wage inequality. *American Economic Journal: Applied Economics*, 5(1):65–103.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

OECD (2010). *OECD Employment Outlook 2010: Moving Beyond the Jobs Crisis*. Organisation for Economic Cooperation and Development, Paris.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics*. Themes in modern econometrics. Cambridge University Press, Cambridge.

Papay, J. P., Willet, J. B., and Murnane, R. J. (2011). Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics*, 161:203–207.

Paul, M. (2016). Is there a causal effect of working part-time on current and future wages? *The Scandinavian Journal of Economics*, 118:494–523.

Picchio, M. and Mussida, C. (2011). Gender wage gap: A semi-parametric approach with sample selection correction. *Labour Economics*, 18:564–578.

Piliavin, J. A. (1987). Age, race, and sex similarity to candidates and voting preference. *Journal of Applied Social Psychology*, 17(4):351–368.

Potrafke, N. (2012). Unemployment, human capital depreciation and pension benefits: An empirical evaluation of German data. *Journal of Pension Economics & Finance*, 11(2):223–241.

Puhani, P. (2000). The Heckman correction for sample selection and its critique. *Journal of Economic Surveys*, 14(1):53–68.

Rahn, W. M. (1993). The role of partisan stereotypes in information processing about political candidates. *American Journal of Political Science*, 37(2):472–496.

Riphahn, R. T. and Schnitzlein, D. D. (2016). Wage mobility in east and west Germany. *Labour Economics*, 39:11–34.

Röttger, C., Weber, B., and Weber, E. (2017). Aktuelle Daten und Indikatoren, Qualifikationsspezifische Arbeitslosenquoten.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2):135–146.

Ruhm, C. J. (1991). Are workers permanently scarred by job displacements? *The American Economic Review*, 81(1):319–324.

Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (German Council of Economic Experts) [SVR] (2014). Jahresgutachten 2014/15: Mehr Vertrauen in Marktprozesse. *Stuttgart: Metzler-Poeschel.*

Sanbonmatsu, K. (2002). Gender stereotypes and vote choice. *American Journal of Political Science*, 46(1):20–34.

Schmieder, J. F., Von Wachter, T., and Bender, S. (2010). The long-term impact of job displacement in Germany during the 1982 recession on earnings, income, and employment. *IAB-Discussion Paper*, No. 1.

Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika*, 66:605–610.

Sigelman, L. and Sigelman, C. K. (1982). Sexism, racism, and ageism in voting behavior: An experimental analysis. *Social Psychology Quarterly*, pages 263–269.

Spitz-Oener, A. (2006). Technical change, job tasks, and rising educational demands: Looking outside the wage structure. *Journal of Labor Econonomics*, 24(2):235–270.

Stambough, S. J. and O'Regan, V. R. (2007). Republican lambs and the democratic pipeline: Partisan differences in the nomination of female gubernatorial candidates. *Politics & Gender*, 3(3):349–368.

Tamm, M. (2013). The impact of a large parental leave benefit reform on the timing of birth around the day of implementation. *Oxford Bulletin of Economics and Statistics*, 7(4):585–601.

Tamm, M., Bachmann, R., and Felder, R. (2017). Erwerbstätigkeit und atypische Beschäftigung im Lebenszyklus–Ein Kohortenvergleich für Deutschland. *Perspektiven der Wirtschaftspolitik*, 18(3):263–285.

Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51:309–317.

Thomas, S. (1991). The impact of women on state legislative policies. *The Journal of Politics*, 53(4):958–976.

Tisch, A. and Tophoven, S. (2012). Employment biographies of the German baby boomer generation. *Schmollers Jahrbuch*, 132(2):205–232.

Vella, F. (1998). Estimating models with sample selection bias: a survey. *Journal of Human Resources*, pages 127–169.

vom Berge, P., Burghardt, A., and Trenkle, S. (2013). Sample of integrated labour market biographies-regional file 1975-2010 (SIAB-R 7510).

Vom Berge, P., Burghardt, A., and Trenkle, S. (2013). Sample of Integrated Labour Market Biographies Regional File 1975-2010 (siab-r 7510). Technical report, FDZ data report, 09/2013, Institute for Employment Research (IAB), Nuremberg.

Weiser, W. R., Levitt, J., Weiss, C., and Overton, S. (2005). Response to the report of the 2005 commission on federal election reform. *Brennan Center for Justice*.

World Bank (2017). Staff estimates based on age/sex distributions of United Nations Population Division's World Population Prospects: 2017 Revision. `https://data.worldbank.org/indicator/SP.POP.TOTL.MA.ZS?locations=DE&year_high_desc=false`. Accessed: 2018-11-13.